

**Einsatz von Kleinrechnerprogrammen
im Bereich nicht-parametrischer Testverfahren
mit Beispielen aus einer empirischen Untersuchung
zur Lage ausländischer Studenten an der
Ruhr-Universität Bochum**

INAUGURAL-DISSERTATION
zur Erlangung des akademischen Grades eines
Doktors der Sozialwissenschaft
der
Ruhr-Universität Bochum
– Abteilung für Sozialwissenschaft –

vorgelegt von
Dipl.-Math.
RENATE PRUST
aus Essen

Bochum 1983

Digitized by the Internet Archive
in 2026

**Einsatz von Kleinrechnerprogrammen
im Bereich nicht-parametrischer Testverfahren
mit Beispielen aus einer empirischen Untersuchung
zur Lage ausländischer Studenten an der
Ruhr-Universität Bochum**

INAUGURAL-DISSERTATION

zur Erlangung des akademischen Grades eines
Doktors der Sozialwissenschaft
der

Ruhr-Universität Bochum

– Abteilung für Sozialwissenschaft –

vorgelegt von

Dipl.-Math.

RENATE PRUST

aus Essen

Bochum 1983

Inhaltsverzeichnis:

	Seite
<u>Kapitel 1: Einleitung</u>	1
1.1 Anlässe der Arbeit	1
1.2 Zielsetzungen	8
1.3 Ausgangssituation	11
<u>Kapitel 2: Programmierung von Mikrocomputern</u>	14
2.1 Programmiersprache	14
2.2 Datenspeicherung	23
<u>Exkurs: Speicherung numerischer Werte als Strings</u>	31
2.3 Dateneingabe	36
2.4 Datenkorrektur	40
<u>Kapitel 3: Methoden der induktiven Statistik</u>	44
3.1 Aufgaben der induktiven Statistik	44
3.1.1 Signifikanztests	44
3.1.2 Parametrische und nicht-parametrische Tests	49
3.2 Prüfung einer empirischen Verteilung auf Abweichung von einer definierten Verteilungsform	52
Binomialtest	52
χ^2 -Anpassungstest	55
Kolmogoroff/Smirnow-Anpassungstest	59
3.3 Prüfung auf Zufälligkeit - Iterationstest	62
3.4 Prüfung von Unterschieden zwischen unabhängigen Stichproben	65
3.4.1 Anpassung der Verteilungen	66
Fisher-Test	66

χ^2 -Homogenitätstest für zwei und mehr Stichproben	68
Wald/Wolfowitz-Runtest	71
Kolmogoroff/Smirnow-Anpassungstest für zwei Stichproben	74
3.4.2 Mittelwertunterschiede	76
Mediantest für zwei und mehr Stichproben	76
Mann/Whitney-U-Test	79
Kruskal/Wallis-H-Test	83
t-Test für zwei unabhängige Stichproben	86
Randomisierungstest	92
Einfaktorielle Varianzanalyse	94
3.4.3 Streuungsunterschiede	97
Varianzquotiententest	98
Bartlett-Test	100
3.5 Prüfung von Unterschieden zwischen verbundenen Stichproben	102
3.5.1 Anteilswertdifferenzen	102
McNemar-Test	103
Cochran-Q-Test	105
3.5.2 Mittelwertdifferenzen	107
Dixon/Mood-Vorzeichen-Test	107
Wilcoxon-Vorzeichen-Rangtest	110
Friedman-Test	112
t-Test für zwei verbundene Stichproben	115
3.6 Prüfung des Zusammenhangs von Merkmalen	117
χ^2 -Unabhängigkeitstest	117
Rangkorrelationskoeffizient nach Spearman	119
Korrelationskoeffizient von Bravais/Pearson	122
3.7 Zusammenfassende Übersicht	125

<u>Kapitel 4: Programmbeschreibung von REP-TEST</u>	129
4.1 Hauptmenu	130
4.2 Programm "TESTVERFAHREN" - Direkte Auswahl der Tests	131
4.2.1 Initialisierung und Menu	132
4.2.2 Unterprogramme	135
4.2.3 Testverfahren	153
4.2.3.1 Binomialtest	154
4.2.3.2 χ^2 -Anpassungstest	155
4.2.3.3 Kolmogoroff/Smirnow-1 Stichproben-Test	157
4.2.3.4 χ^2 -Homogenitätstest	163
4.2.3.5 Fisher-Test	164
4.2.3.6 Kolmogoroff/Smirnow-2 Stichproben-Test	165
4.2.3.7 Wald/Wolfowitz-Runtest	167
4.2.3.8 McNemar-Test	172
4.2.3.9 Cochran-Q-Test	172
4.2.3.10 Runtest	174
4.2.3.11 χ^2 -Unabhängigkeitstest	175
4.2.3.12 Spearman	176
4.2.3.13 Bravais/Pearson	177
4.2.3.14 Mediantest	178
4.2.3.15 Mann/Whitney-U-Test	180
4.2.3.16 Kruskal/Wallis-H-Test	181
4.2.3.17 Dixon/Mood-Vorzeichentest	181
4.2.3.18 Wilcoxon-Test für verbundene Paare	182
4.2.3.19 Friedman-Test	183
4.3 Programm "TESTS MIT HILFEN" - Auswahl der Tests durch das Programm	185
4.3.1 Initialisierung und Fehlerrouitinen	186
4.3.2 Unterprogramm	186
4.3.3 Auswahl eines geeigneten Tests	188
4.3.4 Testverfahren	198

	Seite
<u>Kapitel 5: Beispiele</u>	200
<u>Kapitel 6: Schlußbemerkungen</u>	219
<u>Literaturverzeichnis</u>	222
<u>Anhang</u>	228
I. Speicherplätze und Unterprogramme des APPLE II	228
II. Inhaltsverzeichnis der Systemdiskette	230
III. Programm Listings	231
IV. Fragebogen	286
V. Codeplan	290

Kapitel 1: Einleitung

1.1 Anlässe der Arbeit

Da die Mikrocomputer in immer mehr Bereiche eindringen, ist die Beschäftigung mit den Möglichkeiten dieser Geräte zumindest für jeden, der sich mit elektronischer Datenverarbeitung beschäftigt, unumgänglich. Dabei ist die Abgrenzung der Gruppe der Mikrocomputer etwas schwierig, da die Bezeichnungsweise nicht einheitlich ist. Mikrocomputer, Kleinrechner, Minicomputer, Tischrechner, Personal Computer - alle diese Namen werden in der Literatur und der Fachpresse manchmal synonym benutzt, manchmal jedoch auch als Unterscheidungsmerkmal für unterschiedlich große Arbeitsspeicher eingesetzt. In dieser Arbeit werden die Begriffe Mikrocomputer und Kleinrechner gleichbedeutend benutzt für Rechner, die einen mindestens 40 Zeichen breiten Bildschirm besitzen, eine schreibmaschinenähnliche Tastatur und einen Arbeitsspeicher von 16 bis 64 Kbyte RAM (Random Access Memory). Die Bezeichnung RAM zur Unterscheidung von ROM (Read Only Memory) ist etwas irreführend; denn der Unterschied liegt nicht im wahlfreien bzw. nicht-wahlfreien, d. h. sequentiellen, Zugriff. ROM's sind Speicher, deren Inhalt nur gelesen, in die aber nicht geschrieben werden kann. RAM's erlauben hingegen beide Prozesse, d. h. sie müßten eigentlich Read und Write Memory heißen, doch haben sich die beiden Bezeichnungen RAM und ROM durchgesetzt.

Die Dezentralisierung von Rechenzentren, z. B. über "intelligente" Terminals wie im Rechenzentrum der Ruhruniversität, ist ein sicheres Anzeichen dafür, daß sich Kleincomputer auch in den klassischen Einsatzbereichen von Großrechenanlagen durchsetzen können. Dadurch werden die Großrechner entlastet von der Auswertung kleinerer Datenbestände über kurze, einfache Programme und können ausschließlich genutzt werden für Aufgaben, deren Bearbeitung tatsächlich die großen Kapazitäten erfordert, die die Großrechner zur Verfügung stellen. Die technische Entwicklung hat im Bereich Mikrocomputer Preisensenkungen möglich gemacht, die eine Erschließung immer neuer "Absatzquellen" zur Folge hatten.

Dazu gehören Handwerksbetriebe, kleinere Handelsfirmen, also Betriebe einer Größenordnung, für die die früheren größeren Rechner unrentabel waren, hingegen ein Kleinrechner Vorteile bringen kann, die mit nicht sehr hohen Kosten erreicht werden können. Auch Schulen bedienen sich mittlerweile der neuen Technik, um Verwaltungsaufwand reduzieren zu können - eine Folge auch der Stellenstreichungen -, und als Folge der weiteren Verbreitung von Mikrocomputern werden schon Schüler der Mittelstufe in ihrer Bedienung unterrichtet.

Die Software-Entwicklung verlief jedoch zumindest nicht auf allen Gebieten entsprechend günstig. Insbesondere im Bereich der Statistik gibt es immer noch ein Nachholbedürfnis, und zwar nicht wegen fehlender Programme, sondern wegen des

Mangels an guten Programmen. Denn in vielen Veröffentlichungen offenbart sich ein ausschlaggebender Fehler bei der Programmentwicklung: Es wurde versucht, bestehende Programme und Programmpakete für Großrechenanlagen sozusagen wortgetreu für die Benutzung auf Kleinrechnern zu übertragen. Dabei wird jedoch etwas Entscheidendes übersehen. Mikrocomputer können nicht "im kleinen" genau das leisten, was ein Großrechner leistet.

Der Mikrocomputer kann für viele Probleme ebenso gut eine Lösung liefern, aber nicht unbedingt auf demselben Wege wie der Großrechner.

An dieser Stelle müssen die möglichen Betriebsarten eines Computers angesprochen werden. Man unterscheidet Batch- oder Stapelbetrieb und Dialog- oder interaktiven Betrieb. Der Name Batch bezeichnet ursprünglich den Lochkartenstapel, den der Benutzer im Rechenzentrum abgibt und nach der Bearbeitung durch den Rechner wieder abholt, in der Regel zusammen mit einem Druckerprotokoll. Auf diesen Lochkarten waren sämtliche Befehle an das Betriebssystem des Rechners, das komplette Programm und eventuell Daten enthalten. Im Batch-Betrieb ist also der gesamte Job als ein Block einzulesen und wird dann nacheinander abgearbeitet, ohne daß der Anwender eingreifen kann. Das bedeutet z. B. auch, daß Fehler zum Abbruch des Laufs führen, da "Rückfragen" nicht möglich sind. Dieses Charakteristikum des Batch-Betriebes ist nach der weitgehend vollzogenen Umstellung von Lochkarten- auf Bildschirmeingabe immer noch erhalten.

Im Dialogbetrieb führt der Rechner jeweils nur eine Anweisung aus, danach präsentiert er gegebenenfalls das Ergebnis auf dem Bildschirm und wartet auf eine neue Anweisung. Bei Fehlern, die er als solche erkennen konnte, bricht er ab und zeigt dies durch eine Fehlermeldung an. Der Benutzer kann also nach jeder Anweisung, die unter Umständen allerdings die Bearbeitung eines kompletten Programms enthalten kann, eingreifen. Dadurch entfällt die Notwendigkeit, die gesamte Liste der Anweisungen im Arbeitsspeicher des Rechners bereitzuhalten, wie dies im Stapelbetrieb der Fall ist.

Die Möglichkeiten des Dialogbetriebes müssen bei der Arbeit mit Kleinrechnern in viel stärkerem Maße eingesetzt werden, als dies auf dem Großrechner geschieht und geschehen sollte.

Läßt man Kleinrechner im Batch-Betrieb arbeiten, so verzichtet man auf einen wesentlichen Vorteil der auch "Personal Computer" genannten Geräte. Jeder, der schon einmal während der Hauptnutzungszeit eines Großrechners auf seine Ergebnisse längere Zeit warten mußte, wünscht sich einen Personal Computer, auf dem man "eben mal" ein kleines Programm laufen lassen kann - ohne auf Rechenzeit warten zu müssen, wie es immer wieder beim Großrechner vorkommt, und ohne auf Druckerprotokolle warten zu müssen, wie es beim Großrechner die Regel ist.

Der interaktive Betrieb ist also die Stärke der Mikrocomputer, deren Nutzen man nicht durch ungeschickte, weil großrechnerorientierte Programmierung mindern sollte.

Ein weiterer Punkt, weshalb das einfache Übertragen von Programmen auf den Mikrocomputer meist keine gelungenen Ergebnisse bringt, ist der angesprochene "Kundenkreis". Die Benutzer von Groß-

rechenanlagen bilden eine wesentlich kleinere Gruppe als die der Kleinrechnernutzer, die zudem noch völlig anders geschult sind. Sie haben unter Umständen keinerlei Kenntnisse und Erfahrungen in der elektronischen Datenverarbeitung, geschweige denn im Programmieren. Soweit es sich um den Bereich der statistischen Methoden handelt, die in dieser Arbeit ja angesprochen werden sollen, muß man zudem oft davon ausgehen, daß auch das Wissen über diese Methoden stark eingeschränkt ist. In vielen Fachgebieten hat man schon seit längerem erkannt, daß Statistik nicht nur als mathematische Statistik betrieben werden kann, sondern darüberhinaus Methoden bereitstellt zur Informationsgewinnung, die auch von Nicht-Mathematikern durchaus nutzbringend eingesetzt werden können.¹⁾

Da nun immer häufiger unterschiedlich ausgebildete Personen versuchten, statistische Methoden zu nutzen, wurden Programmpakete entwickelt wie z. B. SPSS (Statistical Package for the Social Science) und BMDP (Bio-Medical Department Programs), die denjenigen die Anwendung statistischer Methoden auf größere Datenbestände ermöglichten, die nicht in der Lage waren, selbst Programme zur Auswertung zu schreiben. Statistische Auswertungen ohne EDV-Anlage sind heute kaum noch denkbar und werden nur in seltenen Fällen noch durchgeführt, z. B. bei sehr kleinen Fallzahlen; auf die Zeitersparnis, die die elektronische Datenverar-

1) Die Ausdehnung der Anwendungsbereiche von statistischen Methoden hat im übrigen auch auf die Entwicklung neuer Verfahren und die Weiterentwicklung alter einen fruchtbaren Einfluß ausgeübt.

arbeitung bewirkt, will kaum noch jemand verzichten.

Auch für Kleinrechner wurden nach und nach immer mehr solcher Pakete entwickelt, z. B. MICRO-SPSS von ROSE (1982) oder K-MICRO-STAT-1 von KLEITER (1981), deren Nennung hier keine Bewertung darstellt.

Ein Nebeneffekt wurde jedoch hervorgerufen durch die Anwender der Programmpakete, die wahllos alles ausnutzten, was machbar war. Wer z. B. eine bivariate Häufigkeitstabelle zwischen den Variablen "Fallnummer" und "Geschlecht" erstellen läßt,¹⁾ hat offensichtlich nicht nachgedacht über sein Vorgehen.

Dieses Phänomen - erst einmal einiges machen, dann weitersehen - kann zwei Gründe haben: Unfähigkeit oder Bequemlichkeit. Die Programmpakete bieten meist ein reichhaltiges Angebot von univariaten, bivariaten und multivariaten Verfahren und Kennzahlen, aus dem auszuwählen natürlich Zeit und Wissen erfordert. In SPSS gibt es z. B. für viele Prozeduren das verlockende Schlüsselwort ALL, das bewirkt, daß die zuvor genannte Prozedur für alle Variablen durchgeführt wird.

Durch die Nutzung solcher Möglichkeiten kann die Erstellung eines SPSS-Laufes schneller und einfacher erfolgen, was allerdings nur eine Verschiebung der Anstrengungen mit sich bringt, da sie bei der nachfolgenden Interpretation der Druckerprotokolle ohnehin geleistet werden müssen. Die Auswahl der für den Einzelfall geeigneten

1) Auch solche extremen Beispiele finden sich in der Raritätensammlung des Rechenzentrums der RUB.

Verfahren aus dem übergroßen Angebot vieler Pakete erfordert fundiertes Wissen über statistische Methoden, deren Anwendungsvoraussetzungen und ihre Umsetzung innerhalb der Programmpakete - und überfordert offensichtlich eine große Gruppe von Benutzern, die nicht in der Lage ist bzw. gewillt ist (schon aus zeitlichen Gründen), sich eingehender mit diesen Problemen zu beschäftigen. Je größer der Kreis potentieller Benutzer von Statistik-Programmpaketen wird, um so größer wird vermutlich die Gruppe der "ungeschulten" Anwender, d. h. das Problem der ungenügenden Vorkenntnisse wird sich bei Benutzern von Kleinrechnern möglicherweise in noch viel stärkerem Maße stellen. Die Statistiker sollten deshalb versuchen, bei der Entwicklung geeigneter Programme die Nutzung statistischer Methoden durch Strukturierung des Problemkreises auch solchen "Dilettanten" zu ermöglichen.

Muß sich denn der gelegentliche Nutzer dieser Methoden tatsächlich in alle Details einarbeiten? Die Frage stellt sich insbesondere für den Bereich der induktiven Statistik. Wahrscheinlichkeitstheorie und statistische Testtheorie sind - auch mathematisch - anspruchsvolle Gebiete, deren gründliche Kenntnis von niemandem ernsthaft erwartet werden kann, der ab und zu eine Hypothesenprüfung durchführen möchte.

Deshalb soll in dieser Arbeit der Versuch unternommen werden, durch Strukturierung des Bereichs der statistischen Analysemethoden - speziell der statistischen Signifikanztestverfahren -, durch Entwicklung geeigneter Programme und durch eine anwendungsbezogene Analyse der Einsatzbedingungen

der verschiedenen Verfahren die Grundlage zu schaffen für eine "neue Sicht" der inferenzstatistischen Verfahren.

Daß dabei zusätzlichen Benutzerkreisen Hilfe gestellt geboten wird, ist ein erfreulicher Begleiteffekt.

1.2 Zielsetzungen

Es soll im folgenden versucht werden, die in 1.1 genannten Aspekte miteinander zu verbinden. Aufgrund des allgemeinen aktuellen Interesses an Mikrocomputern und ihren Einsatzmöglichkeiten soll ein Nachholbedürfnis an speziell für Kleinrechner entwickelten Programmen befriedigt werden, und zwar für den Bereich der induktiven Statistik. Damit auch Nicht-Fachleute das richtige Testverfahren für ihren besonderen Anwendungsfall einsetzen können ohne langwierige Studien der Methoden der Inferenzstatistik, soll ein Verfahren entwickelt werden, das in benutzerfreundlicher Weise schrittweise auf ein geeignetes Testverfahren hinleitet. Die Auswahl eines solchen Tests erfordert die Überprüfung mehrerer Voraussetzungen, z. B. Meßniveau, Anzahl und Größe der Stichproben, Abhängigkeit oder Unabhängigkeit, und natürlich Kenntnis über die zu prüfende Hypothese. Das entwickelte Programmsystem REP-TEST bietet dem Benutzer im interaktiven Betrieb Entscheidungshilfe bei der Klärung dieser notwendigen Fragen, wählt dann entsprechend der gemachten Eingaben das geeignete Testverfahren aus und führt es für den gewünschten Datenbestand

durch. Besonders wichtig bei einem so organisierten Programmsystem ist die Benutzerfreundlichkeit, d. h. ein geringer Schwierigkeitsgrad bei der Anforderung von Eingaben, und die Absicherung gegen Benutzerfehler. Deshalb mußte sich das Hauptinteresse bei der Programmierung darauf richten, die "Systemsprache" so einfach und eindeutig wie möglich zu gestalten. Aus diesem Grund wurde dem Benutzer die Mühe abgenommen, selbst Eingaben zu formulieren; stattdessen wurde die "Menutechnik" verwendet, die nur eine Auswahl zwischen vorgegebenen Alternativen erlaubt. Die Möglichkeiten für Syntaxfehler wurden dadurch erheblich eingeschränkt, was besonders für Anwender ohne EDV-Erfahrung von großem Vorteil ist, da sie nach einer ganz kurzen Einarbeitungszeit das System benutzen können.

Für den Statistikfachmann aber, der einen Kleinrechner einsetzen will, wäre hingegen das "Vortasten" über mehrere Menüs bis hin zum geeigneten Test ein zu langatmiges Verfahren, da er höchstwahrscheinlich schon vorher weiß, welchen Test er einsetzen kann und will. Deshalb wurde für diese Gruppe von Anwendern die zusätzliche Möglichkeit geschaffen, den gewünschten Test direkt anzuwählen.

So bietet dann das vorgestellte Programmsystem REP-TEST sowohl für den Statistiker als auch den Nicht-Statistiker die Möglichkeit, für eine Reihe verschiedener Hypothesen unter verschiedenen Voraussetzungen einen geeigneten statistischen Test mit Hilfe eines Mikrocomputers durchzuführen. Für die vorliegende Arbeit wurde ein APPLE II mit

48 Kbyte Arbeitsspeicher und einem 5 1/4"- Diskettenlaufwerk benutzt. In dieselbe Klasse gehören die ebenfalls weit verbreiteten Rechner von COMMODORE und TANDY TRS-80.

Ein wichtiger Punkt bei der Arbeit mit Kleinrechnern, der hier noch kurz angesprochen werden soll, ist die Dateneingabe und die Datenspeicherung - ein Punkt, über den oft zu leichtfertig hinweggegangen wird. Er kann hingegen als durchaus wesentlich für die praktische Arbeit angesehen werden, da die Dateneingabe oft die "ertragreichste" Quelle für Fehler darstellt. Hierzu waren keine eigenen Entwicklungen erforderlich, da ein ausgezeichnetes Verfahren bereits vorliegt,¹⁾ das in Kap. 2 näher beschrieben wird. Da es ebenfalls für den APPLE II entwickelt wurde, läßt es sich in das System REP-TEST ohne Schwierigkeiten einbauen, ebenso wie die Möglichkeiten der Datenmodifikation, die dieses System DATAN II bietet. Es ist anfängerorientiert, stellt aber doch im ganzen höhere Anforderungen an die Vorkenntnisse des Benutzers, ähnlich wie die bekannten Programmpakete, z. B. SPSS.

Eine Beschreibung der hier übernommenen Teile des Systems ist deshalb in das Kapitel 2 aufgenommen worden.

1) Schnell, Rainer: Datenanalyse mit Kleinrechnern. Entwicklung eines interaktiven Datenanalysesystems für Microcomputer, unveröffentlicht.

1.3 Statistische Tests in Programmpaketen - Ausgangssituation

Anhand des am häufigsten in den Sozialwissenschaften benutzten Programmpaketes SPSS soll erläutert werden, welche Anforderungen die Nutzung von SPSS an den Anwender stellt, um die Notwendigkeit des vorliegenden Systems deutlich zu machen. Zugrundegelegt wurde die Version 7 von SPSS, obwohl bereits auf vielen Großrechnern Version 8 bzw. sogar Version 9 implementiert ist, die aber im Hinblick auf das Problem der Auswahl des geeignetsten Tests keine Änderungen enthalten.

In der deutschsprachigen Programmbeschreibung zu SPSS 7 findet sich gleich zu Beginn des Abschnitts NPAR TESTS ein Hinweis auf das größte Problem vieler Anwender:

"In diesem Zusammenhang sei angemerkt, daß nicht etwa das SPSS-Programm die einzelnen Variablen auf Adäquatheit bezüglich des vorgesehenen Signifikanztests prüft."¹⁾ Auf den insgesamt 14 Seiten muß für alle möglichen Tests der Prozedur NPAR TESTS der Aufbau der SPSS-Prozedurkarte im einzelnen beschrieben werden. Sich durch diese Fülle von Möglichkeiten, eine einzelne SPSS-Anweisung zu schreiben, durchzuarbeiten, ist allein schon ein Pensum, daß viele zur Oberflächlichkeit verführen wird. Dazu kommen dann noch eine Reihe von Optionen, die angefordert werden können. Auch ein versierter Statistiker muß da zunächst einmal

1) Beutel, P. u. a.: SPSS 7, 2. Auflage, Stuttgart/New York, 1978, S. 134

suchen, bis er die richtige Form des SPSS-Befehls und die passende Option gefunden hat - zumindest, wenn er nicht sehr häufig die Prozedur NPAR TESTS abrufen.

Ein etwas ungeübter Anwender soll nun ja auch noch nach Skalenniveau, Anzahl der Stichproben usw. das richtige Testverfahren aussuchen.

Wenn er weiß, daß seine Daten Ordinaldaten sind und daß er zwei unabhängige Stichproben vorliegen hat (dies ist ja noch leicht zu übersehen), nimmt er nun zur Prüfung von Unterschieden den Kolmogorov-Smirnov-Test oder den Wald-Wolfowitz-Test?¹⁾ Beide sind im Angebot der SPSS-Prozedur vorhanden (insgesamt zeigt die Übersicht in SPSS 7, S. 135, 24 Tests). Spätestens diese Entscheidung überfordert viele Anwender und sie wählen deshalb den unerwünschten Weg, alles Denkbare auch einmal vom Rechner ausarbeiten zu lassen.

Neben dem Punkt des vergrößerten, vor allem auch veränderten Anwenderkreises findet sich hier ein weiterer Aspekt, der für Kleinrechner ein System mit "Vorauswahl" sinnvoll erscheinen läßt. Beim Großrechner fällt vielleicht ein Test mehr oder weniger nicht so sehr ins Gewicht hinsichtlich der Rechenzeit, beim Kleinrechner jedoch, der für die einzelne Prozedur wesentlich mehr Zeit benötigt, muß das Vermeiden unnötiger Berechnungen immer im Auge gehalten werden.

1) Zu den Tests vergl. Kap. 3

Als Beispiel für Fehlermöglichkeiten sei einmal die Prozedurkarte für Signifikanztests für eine Stichprobe ausgewählt; sie hat allgemein die Form:

1	16
NPAR TESTS	BINOMIAL oder CHI-SQUARE oder K-S oder RUNS [(Parameter)] = Variablenliste [Parameter] [/nächster Signifikanztest...]

Nach der Entscheidung über das richtige Verfahren ist also immer noch die Frage offen, ob und welche Parameter angegeben werden sollen oder müssen.

Hier ist nun wirklich Abhilfe zu schaffen, wie es insbesondere in einem interaktiven System durch mehr Benutzerfreundlichkeit geleistet werden kann - durch detaillierte Strukturierung des Bereichs, sodaß zumindest Erläuterungen der geforderten Eingaben in Kommunikation mit dem System angeboten werden. Gerade wenn nicht, wie im Batch-Betrieb, alle notwendigen Informationen an den Rechner vorab eingegeben werden müssen, sollte "schrittweises" Arbeiten möglich (und wünschenswert) sein. Auch die bereits veröffentlichten Programmpakete für Kleinrechner bieten - genau wie SPSS - keine Entscheidungshilfen, sondern überlassen den Anwender bei der Auswahl des Tests sich selbst. Ein entscheidender Nachteil, ganz abgesehen von teilweise erheblichen Mängeln in der Programmierung¹⁾ und der Verwendung des Batch-Betriebes.

1) Die sich z. B. in sehr langen Rechenzeiten niederschlagen, wie z. B. bei TASSEL, D. V.: BASIC-Pack Statistics, Programs for Small Computers, Englewood Cliffs 1981

Kapitel 2: Programmierung von Mikrocomputern

2.1 Programmiersprache

Die für Mikrocomputer am meisten verbreitete Sprache ist BASIC (Beginners All-Purpose Symbolic Instruction Code), das 1965 von J. KEMENY und T. KURTZ entwickelt wurde. Wie der Name sagt, ist es eine für Programmieranfänger entwickelte höhere Programmiersprache für alle Anwendungen, d. h. sie ist nicht im Hinblick auf die Lösung eines konkreten Anwendungsgebietes entstanden, wie dies zum Beispiel bei COBOL (Common Business Oriented Language) oder FORTRAN (Formula Trans-formation) der Fall war.

Es gibt für viele Kleinrechner inzwischen auch Compiler für andere Programmiersprachen (für den APPLE z. B. COBOL-, FORTRAN-, PASCAL-, ADA-, FORTH- und LISP-Compiler), doch ist BASIC immer noch die Standardsprache der Mikrocomputer, die zumeist mit einem BASIC-Interpreter statt eines Compilers arbeiten. Ein Compiler ist ein Übersetzungsprogramm, das Programme, die in einer höheren Programmiersprache geschrieben sind, in Maschinensprache übersetzt. Ein Interpreter hingegen "interpretiert" gewisse Zeichenfolgen als Befehle der jeweiligen Programmiersprache und übersetzt den gesamten Befehl (also nicht zeichenweise wie der Compiler) in einen einzigen Zahlen-code. Jedes BASIC-Befehlswort benötigt so nur 1 byte Speicherplatz unabhängig von seiner ursprünglichen Länge. Der Grund für die weitere Verbreitung von BASIC-Interpretern ist in dem wesent-

lich geringeren benötigten Speicherplatz zu sehen; hingegen arbeiten Compiler schneller. Da der Einsatz anderer Sprachen meist eine Erweiterung des Arbeitsspeichers voraussetzt, ist die Verwendung von BASIC leicht erklärlich. Die leichte Erlernbarkeit von BASIC ist unbestritten, der einfache Aufbau dieser Programmiersprache ermöglicht dem Anfänger relativ schnellen Zugang zum Programmieren; einfache eigene Programme können nach kurzer Zeit selbst erstellt werden. Dadurch eignet sich BASIC ausgezeichnet für Lehrveranstaltungen. Für komplexe Programmsysteme erfordert die Programmierung in BASIC schon ein hohes Maß an Disziplin in der Programmieretechnik, da Möglichkeiten zur strukturierten Programmierung weitgehend fehlen. Damit ist die hierarchische Strukturierung eines Programms in mehrere Stufen gemeint, bis hinunter zu den einzelnen Rechenanweisungsblöcken. Blöcke der gleichen Stufe können problemlos ausgetauscht werden, falls Programmänderungen erforderlich sind, wegen ihrer Unabhängigkeit, da Datenaustausch nur auf der jeweils höheren Hierarchiestufe erfolgt.

APPLESOFT ist einer von vielen BASIC-Dialekten, die Erweiterungen zum ursprünglichen BASIC liefern. Durch ihre große Anzahl sind jedoch BASIC-Programme nicht mehr auf jeder beliebigen Maschine einsetzbar, es müssen oft Übersetzungen von einem Dialekt in den anderen geschrieben werden, die aber in der Regel kein Problem darstellen, sofern nicht mit maschinenabhängigen Unterprogrammen gearbeitet wird (z. B. CALL-868, siehe dazu Anhang) bzw. Speicheradressen benötigt werden,

die ebenfalls maschinenabhängig sind. Zu den Vorteilen von APPLESOFT gehört die TRACE-Funktion zum Debugging, d. h. zur Fehlersuche. Mit dem Kommando TRACE kann die "Spur" der Rechnerarbeit verfolgt werden. Auf dem Bildschirm erscheint immer die Zeilennummer der soeben bearbeiteten Anweisung. Durch die Folge dieser Zeilennummern können z. B. falsche Sprünge oder Verzweigungen eines Programms entdeckt werden.

Das DOS 3.3 (Disk Operating System) bietet mit der MON-Funktion eine Möglichkeit, Fehler in Eingabedaten, Ausgabedaten oder DOS-Anweisungen zu erkennen. Ist diese Funktion in Kraft, so werden die auf Diskette oder von Diskette übertragenen Informationen auf dem Bildschirm dargestellt, und zwar Informationen aller 3 Gruppen oder nur ausgewählter Gruppen - je nach Spezifizierung des Kommandos MON.

Diese beiden Debugging-Kommandos bleiben solange in Kraft, bis sie aufgehoben werden, gleich welche Kommandos anschließend eingegeben werden.

Der EXEC-Befehl des DOS ermöglicht es, in Textdateien gespeicherte Anweisungen (Kommandos oder Programmzeilen) zeilenweise in den Arbeitsspeicher zu übertragen. Kommandos werden direkt ausgeführt (wie bei Eingabe über die Tastatur), Programmzeilen nach erfolgtem Kommando RUN. Es kann also z. B. ein komplettes Programm zusammen mit RUN als Textdatei abgespeichert werden auf Diskette, dann wird durch EXEC, gefolgt von einem Dateinamen, das Programm eingelesen und gestartet. Hervorzuheben ist ebenfalls die Möglichkeit, im ESC-Modus (ESC = Escape) Fehler im

Programm in einfachster Weise korrigieren zu können. In diesem Modus haben einige Buchstaben-Tasten Cursorsteuerfunktion, d. h. sie steuern horizontale und vertikale Bewegungen des Lichtzeichens, das die Bildschirmposition des nächsten eingegebenen Zeichens anzeigt.

Als nachteilig anzusehen ist die Tatsache, daß der Interpreter bei jedem Systemstart neu geladen werden muß und einen Teil des 48 K RAM wegnimmt (etwa 10 K laut APPLE II Anwenderhandbuch). Für die Gestaltung des Bildschirms muß bei APPLESOFT viel Programmierarbeit aufgewandt werden, wie bei vielen anderen Rechnern dieser Klasse auch, da ein PRINT USING-Befehl zur formatierten Bildschirmanzeige fehlt. Zumindest sind aber die Funktionen HTAB und VTAB - auch mit variablem Funktionswert - implementiert, durch die der Cursor leicht an eine bestimmte Bildschirmstelle gebracht werden kann. Das Fehlen von WHILE ... WEND - bzw. REPEAT ... UNTIL - Statements ist ein Nachteil, der die Programmier-technik vom Ideal der strukturierten Programmierung stark unterscheidet, der jedoch ebenfalls vielen Rechnern der Klasse gemein ist.¹⁾

Trotz der genannten Nachteile sind BASIC im allgemeinen und APPLESOFT im besonderen durchaus brauchbar für die Programmierung auch von komplexen Problemen. Die genannten Strukturen WHILE ... WEND und REPEAT ... UNTIL können auch simuliert werden²⁾ und der fehlende PRINT USING-

1) Eine Ausnahme bildet z. B. der alphasoft von Triumph-Adler, der aber mit dem CP/M-Betriebssystem arbeitet. Zudem liegt er im Preis etwas höher.

2) MAIORANA, J. A.: Changing a BASIC FOR ... NEXT Loop into a REPEAT ... UNTIL Loop, in: BYTE 1981/9 S. 162

Befehl ist durch geschickten Einsatz der zahlreich vorhandenen Stringmanipulationsmöglichkeiten mit etwas Aufwand zu ersetzen.

Bei der Arbeit mit Mikrocomputern ergeben sich zwei Hauptprobleme, die sich für Großrechenanlagen nicht oder zumindest nicht in demselben Maße stellen: die geringe Speichergröße und die geringe Ausführungsgeschwindigkeit.

Die geringe Speichergröße wirkt sich insbesondere auf die Bereitstellung der Daten für die Auswertung aus. Dazu soll in den nächsten Abschnitten ausführlich über Dateneingabe, Datenspeicherung und Datenmodifikation gesprochen werden.

Sofern die Ausführungsgeschwindigkeit nicht von der Maschine abhängt (d. h. Taktfrequenz und Wortlänge) bzw. vom Compiler oder Interpreter - diese beiden Faktoren sind in der Regel durch den Programmierer nicht zu beeinflussen - ist die Struktur des Programms der Hauptansatzpunkt, um kürzere Ausführungszeiten zu erreichen. Dazu gehören zum einen die verwendeten Algorithmen, d. h. die Wege, die beschritten werden zur Lösung eines Problems. Wie groß bei unterschiedlichen Algorithmen für dasselbe Problem die Unterschiede in den Ausführungszeiten sein können, ist der Arbeit von SCHNELL (1982) zu entnehmen, der für die Erstellung des DATAN II-Datenanalysesystems Vergleichsmessungen durchführte für Sort, Ranksort und Rank-Algorithmen, und zwar speziell für den APPLE II. Sort-Algorithmen dienen dazu, eine Reihe von numerischen Werten der Größe nach (meist in aufsteigender Folge) anzuordnen. Dabei müssen Kleiner- oder Größer-Vergleiche durchgeführt werden, bei N Zahlen im

direkten Vergleich - jede Zahl wird mit allen nachfolgenden verglichen - sind $N(N-1)/2$ Vergleiche nötig.

Ranksort-Algorithmen leisten neben dem Sortieren der Werte einer Variablen X noch das Vertauschen der entsprechenden Werte einer Variablen Y, wenn es sich z. B. um Merkmalsausprägungen zweier Variablen bei derselben Person handelt.

Rank-Algorithmen ersetzen die Werte in einer sortierten Reihe durch Rangplatzzahlen, wobei gleichen Ursprungswerten die gleiche "durchschnittliche" Rangplatzzahl zugewiesen werden muß.

Bei den Vergleichsmessungen stellte sich auch heraus, daß der für Kleinrechner geeignetste Algorithmus nicht immer derjenige ist, der auf Großrechenanlagen die schnellste Ausführungsgeschwindigkeit erreicht. So schied z. B. der Rank-Algorithmus AS26 von FREEMANN¹⁾ aus, da er mit 4 geschachtelten FOR ... NEXT-Schleifen arbeitet, was für Kleinrechner untragbar ist.

Wegen der vielen Testläufe, die SCHNELL auf dem APPLE II durchführte, soll hier auf eine ausführliche Behandlung dieser drei Algorithmengruppen verzichtet werden. Entsprechend seiner Argumentation werden für die genannten Probleme folgende Algorithmen ausgewählt.

Für das Sortieren eines numerischen Arrays, d. h. eines Feldes, dessen Werte unter demselben Namen durch die Angabe der Feldposition angesprochen

1) FREEMANN, P. R.: Algorithm AS 26. Ranking an Array of Numbers, In: Applied Statistics 1974/19, S. 111-113

werden können, im direkt adressierbaren Speicher wurde ein modifizierter QUICKSORT-Algorithmus von KOLBE¹⁾ ausgewählt, der für ein Array von 200 Elementen 24 Sekunden benötigt (auf dem APPLE).

Nach einer kleinen Fehlerkorrektur schien für das Problem Ranksort - d. h. ein Array X wird sortiert und die Werte eines Arrays Y werden mitgeführt - das Programm von VELLEMAN und HOAGLIN²⁾ am besten geeignet.

Da für einige Tests auch Rangplätze benötigt werden, braucht man ein Unterprogramm, das Werte eines Arrays in Ränge transformiert. Die bisher vorhandenen benötigten wegen der verwendeten umständlichen Algorithmen so lange Ausführungszeiten, daß SCHNELL eine eigene Ranking-Subroutine schrieb,³⁾ die für diese Arbeit noch einmal verbessert werden konnte. Das Unterprogramm berechnet auch die für Tests erforderlichen Korrekturgrößen beim Auftreten von Verbundwerten.⁴⁾

Die Auswahl der richtigen Algorithmen kann ganz entscheidende Auswirkungen auf die für ein Programm benötigte Rechenzeit haben; wie groß der Zeitgewinn ist, richtet sich nach der Größe der

1) KOLBE, W.: Sortieren in Sekundenschnelle, in: CHIP 1982/1, S. 9

2) VELLEMAN, P. F./HOAGLIN, D. C.: Applications, Basics, and Computing of Exploratory Data Analysis, Boston 1981, S. 304

3) Z. B. benötigte Programm 13 "RANK" von TASSEL 3 min. für ein 200-Elemente-Array (1981)

4) Hierzu vgl. die Beschreibung der Tests in Kapitel 3.

bearbeiteten Arrays und beim Sort-Algorithmus z. B. danach, ob unsortierte oder umgekehrt sortierte Werte im Array enthalten sind.

Außerdem gibt es noch einen zweiten Faktor für die Zeitersparnis bzw. Zeitverschwendung, nämlich die Programmiertechnik, die Art und Weise, wie die Algorithmen in die jeweilige Programmiersprache, also für diese Arbeit BASIC, übertragen werden.

Generell ist zu sagen, daß Sprünge, d. h. GOTO-Befehle, nach Möglichkeit vermieden werden sollten. Jeder Sprung kostet Zeit - je nach Sprungweite, aber mindestens $3 \cdot 10^{-3}$ Sekunden. Und zudem läßt häufiges Hin- und Herspringen im Programm das Programm unübersichtlich, manchmal unlesbar werden. Änderungen von solchen Programmen gestalten sich höchst schwierig und ziehen oft Fehler nach sich, d. h. Software-Wartung kann kaum geleistet werden. Natürlich kann ein sinnvoller Programmaufbau bei gänzlichem Verzicht auf Sprünge nicht zustande kommen, doch muß ihre Zahl rigoros klein gehalten werden.

Konstanten sollten durch Variablen ersetzt werden, da das Umwandeln einer Konstanten gut sechs- bis siebenmal soviel Zeit benötigt wie das Suchen einer Variablen. Das Suchen nach dimensionierten Variablen im Speicher nimmt mehr Zeit in Anspruch als das Suchen einer einfachen.

Eine FOR...NEXT-Schleife ohne Index wird schneller abgearbeitet als eine mit Index.

Solche "Kleinigkeiten" müssen bei Kleinrechnern beachtet werden, damit effizient gearbeitet wer-

den kann. Es gibt noch eine Vielzahl anderer Möglichkeiten, bei einzelnen Anweisungen Zeit zu sparen.

Der bei Kleinrechnern "knappe" Speicherplatz, der insbesondere gründliche Überlegungen über die Datenspeicherung erfordert, ist auch bei der Programmiertechnik zu berücksichtigen. Erstes Gebot ist es, mehrere Anweisungen pro Zeile zu verwenden, da jede Zeilennummer Platz kostet. Dieselben Variablen sollten mehrfach - mit wechselnden Bedeutungen - benutzt werden, um die Gesamtzahl der verwendeten Variablen so klein wie möglich zu halten. Geschickte Abkürzungen für die Variablennamen erleichtern zwar für den ungeübten Leser das Verständnis eines Programms, doch ein solches Verfahren kann bei umfangreichen Programmen schnell dazu führen, daß die Variablen den numerischen Speicher ausschöpfen. Indizierte Variablen benötigen mehr Platz als einfache, allerdings kann dieser Nachteil in vielen Fällen durch die wesentlich platzsparendere Programmierung des gegebenen Problems bei Verwendung von indizierten Variablen ausgeglichen werden bzw. oft kann sogar Platz gespart werden. Die Alternativen müssen im Einzelfall geprüft werden. Wichtig aber ist, daß nicht zu großzügig dimensioniert wird. Bis zu einer Feldgröße von 10 brauchen Arrays in APPLESOFT nicht explizit dimensioniert zu werden. Es ist jedoch sinnvoll, bei kleineren Feldern trotzdem zu dimensionieren, um Platz zu sparen. Leider ist es bei größeren Programmen oft nötig, auf erklärende Bemerkungen im Programm (REM-State-

ments) zu verzichten. Die größere Unübersichtlichkeit muß in Kauf genommen werden, wenn in Frage steht, ob der vorhandene Arbeitsspeicher ausreicht. Dies um so eher bei so ausgabeintensiven Programmen, wie REP-Test es wegen seiner Menutechnik darstellt.

Kurze, häufig aufgerufene Unterprogramme, die von mehreren Programmteilen her angesprochen werden, sollten am Anfang des Programms aufgenommen werden, da sonst die Zeit zum Suchen der genannten Zeile und die Ausführungszeit des Programms in einem zu ungünstigen Verhältnis zueinander stehen. Lange Unterprogramme sollten hingegen ans Ende plziert werden, als jeweils abgeschlossene Blöcke.¹⁾

2.2 Datenspeicherung

Das Problem der Datenspeicherung betrifft 2 Ebenen: sowohl die temporäre als auch die permanente Speicherung müssen befriedigend gestaltet werden. Im Bereich der Statistik wird üblicherweise mit einer Datenmatrix gearbeitet, die pro Fall eine Zeile enthält mit den nacheinander aufgeführten Werten aller Variablen. Eine so gestaltete Datenmatrix in ein Array X (I,J) (I=Fallindex, J=Variablenindex) einzulesen und dieses Feld nun für die Auswertungsprozeduren im Arbeitsspeicher

1) Weitere Hinweise zum Programmaufbau finden sich z. B. bei DAHL, O.-J./DIJKSTRA, E. W./HOARE, C. A. R.: Structured Programming. London/New York 1972, S. 1-82

zur Verfügung zu haben, ist die einfachste Form der Datenbereitstellung. Doch sind Kleinrechner damit überfordert. Die ersten Vorarbeiten für die vorliegende Arbeit wurden durchgeführt auf einem COMMODORE-Rechner CBM 3016 anhand eines Datenbestandes, der 68 Variablen umfaßte mit einer Zeilenlänge von 71 Zeichen pro Fall. Da nur ein Kassettenrecorder als externes Speichergerät für diesen Rechner zur Verfügung stand mit sehr langen Zugriffszeiten auf Dateien, sollte der traditionelle Weg beschritten werden, nämlich: der Datenbestand steht in einem zweidimensionalen Array X für alle Berechnungen zur Verfügung. Schon bei 45 Fällen wäre der gesamte Arbeitsspeicher des CBM ausgeschöpft, denn zur Speicherung eines Arrays im RAM werden zunächst einmal 5 byte benötigt für den Arraynamen (2 byte), für die Anzahl der bytes bis zum Beginn des nächsten Feldes (2 byte), und die Anzahl der Dimensionen (1 byte); hinzu kommen je 2 byte für den größten Index jeder Dimension und 5 byte für jedes Arrayelement. Sei d die Anzahl der Dimensionen, N_i der größte Index der i -ten Dimension ($i=1, \dots, d$), so berechnet sich der für ein Fließkommaarray benötigte Platz SP in byte nach der Formel:

$$SP = 5 + 2 \cdot d + 5 \cdot (N_1 + 1) \cdot \dots \cdot (N_d + 1),$$
 für das obige Beispiel ergibt sich:

$$SP = 5 + 2 \cdot 2 + 5 \cdot 45 \cdot 68 = 5 + 4 + 15300 = 15309 \text{ byte.}^{1)}$$

1) Diese Angaben sind fast allen Anwenderhandbüchern zu entnehmen, wenn auch nicht immer in dieser formalen Schreibweise.

Pro zusätzlichen Fall kommen $5 \cdot 68 = 340$ bytes hinzu, so daß auch der größere Speicher des APPLE II in diesem Beispiel für nur 67 Fälle Platz bietet.¹⁾

Als Alternative bot sich zunächst die Möglichkeit an, für jeden Fall einen String aus allen Variablenwerten einzulesen und abzuspeichern.²⁾

Nachdem sich jedoch die Gelegenheit ergeben hatte, statt des COMMODORE-Rechners den APPLE zu benutzen mit Diskettenlaufwerk, konnten auch andere Alternativen in die Überlegungen einbezogen werden.

Bei der Bildung von Files, d. h. Dateien, stehen grundsätzlich zwei Möglichkeiten zur Verfügung: Rndom-Access-Files oder Sequential Files.

Das Aufsuchen einer bestimmten Information beginnt bei Sequential Files immer am Anfang der Datei; es müssen alle vorher gespeicherten Informationen gelesen werden, bis die gewünschte Information gefunden ist. Dateien mit wahlfreiem Zugriff hingegen sind in gleich lange Records, d. h. Informationseinheiten, aufgeteilt und jede Position innerhalb eines Records kann unmittelbar angesprochen werden, d. h. ohne andere, nicht benötigte Informationen zu lesen. Da die meisten Auswertungsprozeduren jeweils alle Fälle für eine, zwei oder mehr Variablen (oder Stichproben) benötigen, lag es nahe, die übliche Datenmatrix zu transponieren oder anders gesagt: die Werte der Datenmatrix nicht horizontal, sondern vertikal zu je einem File zusammenzufassen. Dies hat folgenden Grund: Wegen des ge-

1) Der gesamte numerische Speicher des APPLE ermöglicht 4600 Datenpunkte zu je 5 bytes.

2) Vgl. dazu den Exkurs ab S. 31.

nannten Platzmangels können sinnvollerweise immer nur die Daten im Arbeitsspeicher zur Verfügung stehen, die für das jeweilige Verfahren benötigt werden. APPLESOFT - DOS gestattet zwar Random-Access-Files, die es ermöglichen würden, auf eine bestimmte Information, ohne eine komplette Variablenliste zu lesen, im Record jedes einzelnen Falles zuzugreifen, doch die Zugriffszeiten sind für interaktives Arbeiten zu lang. Schneller ist deshalb das Speichern in Sequential Files, und zwar in je einem File pro Variable. Jede Diskette des APPLE II kann maximal 49 Variablen zu je 200 Fällen fassen.¹⁾ Die Zugriffszeit beträgt aber immer noch 25 Sekunden für eine Variable. Da auch diese Zeit für ein interaktives System noch zu lang erschien, wurde von SCHNELL (1982) eine andere Methode entwickelt. Sie arbeitet ebenfalls mit je einem Sequential-File pro Variable, der die Variablenwerte aller Fälle enthält. Es können aber maximal 98 Variablen zu je 200 Fällen auf einer Diskette gespeichert werden und die Zugriffszeit verkürzt sich auf 5 Sekunden.²⁾ Mit Erlaubnis des Verfassers wird dieses Verfahren auch für REP-TEST verwendet.

Für die Speicherung maschinensprachlicher Programme bietet der APPLE II die Befehle "BSAVE" und "BLOAD", mit denen Teile des Speichers als binäre bzw. hexadezimale Zahlen gespeichert und

1) Bei mehr Variablen und 200 Fällen muß eine weitere Datendiskette angelegt werden.

2) Bei weniger Fällen ist die Zeit natürlich kürzer.

gelesen werden. Wenn man nun weiß, in welchem Speicherbereich sich die eingegebenen Daten befinden, kann diese Information komplett als maschinensprachliches File auf der Diskette permanent gespeichert werden. Dazu muß man die Startadresse und die Länge der Informationen kennen.

Hat man alle Werte einer Variablen in ein Array des Arbeitsspeichers eingelesen,¹⁾ so muß zunächst der Speicherbereich gefunden werden. Da sich der Beginn des numerischen Speichers mit jeder neudefinierten Variablen verschiebt, kann dies nur über eine Suchschleife erfolgen, die den gesamten numerischen Speicher nach dem gewünschten Array und zwar nach dem Arraynamen durchsucht.²⁾

Vereinfachend wurde für dieses Array immer der Name IO vorgesehen und seine Länge immer auf 201 (200 Datenpunkte und 2 Kontrollgrößen) festgelegt. Eine variable Dimensionierung wäre zwar möglich, würde auch die Zugriffszeit auf 4 Sekunden verkürzen, doch der erforderliche erhöhte Programmieraufwand würde diesen Vorteil wieder zunichte machen. Durch die einmalige Festlegung des Namens IO ist eine Übertragung des IO-Arrays aus dem entsprechenden Hauptarray (oder in das Hauptarray) erforderlich. Der Ablauf der Prozedur sieht wie folgt aus:

1. Zunächst wird das zu speichernde Array in das IO-Array kopiert.

1) Das Verfahren der Dateneingabe wird im folgenden Abschnitt 2.3 beschrieben.

2) Beginn und Ende des numerischen Speichers finden sich in den Speicherstellen 107/108 bzw. 109/110.

2. Das COMPUTE I/O-Modul sucht nun das IO-Array im numerischen Speicher, wobei die Variablen ST% bzw. EN% auf den Beginn bzw. das Ende des numerischen Speichers gesetzt werden. In der Suchschleife wird nach den ASCII-Code-nummern der Buchstaben I und O (73 und 79) sowie des Wertes 1017 im OFFSET-Zeiger (249 und 3) in 4 aufeinanderfolgenden Speicherstellen gesucht. Wird diese Kombination gefunden, so wird der Variablen SI die Startadresse zugewiesen. Wird das Array nicht gefunden, erfolgt eine Fehlermeldung.
3. Mit BSAVE wird nun das IO-Array zusammen mit dem Variablennamen auf Diskette gespeichert.¹⁾

Das Abrufen von Informationen von der Diskette benötigt zu dem COMPUTE I/O-Modul noch ein Lademodul und die Variablenidentifikation.

Die Variablenidentifikationssubroutine fordert zunächst die Eingabe des Namens der Variablen, die geladen werden soll. Nun wird zunächst der VARCATALOG der Datendiskette nach dem eingegebenen Namen durchsucht (dabei führen zusätzliche blanks zu einem neuen Namen). Falls er nicht gefunden wird, erfolgt eine Fehlermeldung und die Eingabe des Namens wird neu verlangt. Ansonsten wird zunächst geprüft, ob die Variable bereits geladen wurde, sich also in der CPU befindet. Ist dies der Fall, endet das Unterprogramm.

1) Nähere Erläuterungen folgen in 2.3

Befindet sich die Variable nicht im Arbeitsspeicher, erfolgt nach einer Prüfung, ob der numerische Speicher weitere Variablen fassen kann (mehr als 15 Variablen mit 200 Fällen können nicht temporär gespeichert werden), der Aufruf des Lademoduls.

Das Lademodul bildet den Namen des Variablenfiles auf der Diskette aus dem Variablennamen, VARDATA FILE und dem Filenamen und ruft das COMPUTE I/O-Modul auf, das - wie oben beschrieben - die Speicherposition des I/O-Arrays bestimmt. Nun werden die Daten in das I/O-Array geladen, dieses wird in das entsprechende Hauptarray kopiert. Zusätzlich erfolgt noch eine Lademeldung. Damit stehen die Daten der für eine Auswertungsprozedur benötigten Variablen im Arbeitsspeicher zur Verfügung. Falls in der Variablenidentifikationssubroutine festgestellt wird, daß schon 15 Variablen geladen sind, muß eine dieser Variablen durch die neue ersetzt werden. Zu diesem Zweck kann ein CPU-Variablenkatalog angefordert werden, aus dem die zu streichende Variable ausgewählt werden kann.

Zusammengenommen stellt dieses Verfahren von SCHNELL einen geglückten Versuch dar, das Problem des "knappen Speichers" für Datenanalysesysteme zu lösen. Es können gleichzeitig 15 Variablen in der CPU zur Verfügung stehen; dies erscheint durchaus ausreichend. Auch die Beschränkung auf 200 Fälle kann kaum als entscheidender Nachteil gesehen werden, da bei noch höherer Zahl von Fällen die Bearbeitung auf dem Kleinrechner ohnehin nicht mehr angemessen erscheint. Spätestens

dann hat der Großrechnereinsatz seine Berechtigung.

Die Tatsache, daß die "Kommunikation" mit der Maschine bei Eingaben, Fehlermeldungen und anderen Bildschirmanzeigen in englischer Sprache erfolgt, ist ohne weiteres hinzunehmen, da auch die Meldungen des Betriebssystems bei Kleinrechnern - fast ohne Ausnahme - in Englisch abgefaßt sind (wenn sie nicht nur aus Abkürzungen oder Fehlercodes bestehen).

Auch das im Exkurs beschriebene Verfahren bietet nicht so viele Vorteile wie das von SCHNELL entwickelte, auch nicht soviel Benutzerkomfort, weshalb für die Programmierung des APPLE II darauf verzichtet wurde. Für andere Rechner, besonders mit Kassettenrecordern gekoppelte kann es allerdings noch zum Einsatz kommen.

Exkurs: Speicherung numerischer Daten als Strings

Die genannten Probleme des geringen Speicherplatzes verschärfen sich, wenn statt eines Diskettenlaufwerks nur ein Kassettenrecorder zur Verfügung steht, da in diesem Fall wegen der geringen Geschwindigkeit eines Recorders die Alternative, jeweils nur einen Fall mit allen Variablen oder eine Variable für alle Fälle einzulesen, ausscheidet. Da potentielle Anwender des Systems REP-TEST unter Umständen nur diese Möglichkeit der Permanentenspeicherung von Daten zur Verfügung haben, soll in diesem Exkurs kurz das Vorgehen erläutert werden. Es kann insbesondere dann Speicherplatz sparen helfen, wenn die Variablen nur ein- oder zweistellige Werte annehmen.

Auf Seite 24 wurde errechnet, daß bei 68 Variablen in der Datenmatrix der benötigte Speicherplatz SP (in byte) eine Funktion der Anzahl der Fälle N ist, nämlich $SP = 5 + 2 \cdot 2 + 5 \cdot N \cdot 68 = 9 + 340 \cdot N$. Wenn die Variablen 1- oder 2-spaltig sind, benötigten ihre Werte, als Zeichenfolge¹⁾ betrachtet, nur jeweils 1 bzw. 2 byte, für jeden Fall im Beispiel gerade 71 byte; wird für jeden Fall ein String kreiert, der alle Zahlen unmittelbar hintereinander aufgeführt enthält - wie dies auch auf Lochkarten der Fall ist -, so kann viel Speicherplatz eingespart werden. Das Array X\$, das alle diese Strings enthält, ist nur noch eindimensional. Ein Stringarray benötigt zunächst einmal genau wie ein Fließkommaarray 5 byte Systeminformation (Name, OFFSET-Zeiger, Anzahl der Dimensionen)

1) Eine solche Zeichenfolge aus Zahlen, Buchstaben oder Symbolen heißt String.

und pro Dimension 2 byte für den größten Index der Dimension. Für jedes Element des Arrays werden weitere 3 byte gebraucht, und zwar 2 für die Speicheradresse, an der der String beginnt, und 1 für die Länge des Strings.

Im Speicherbereich, der die Array-Variablen verwaltet, finden sich also alle Informationen, die nötig sind, um den in einem anderen Bereich des Speichers vorhandenen String abzurufen. Die Variablenwerte selbst, also die Strings, benötigen jeweils pro Zeichen ein byte. Sei wie vorher d die Anzahl der Dimensionen, N_i der größte Index der i -ten Dimension, $i=1, \dots, d$, so ergibt sich für den benötigten Speicherplatz SP einer indizierten Stringvariablen $SP_g = 5 + 2 \cdot d + 3 \cdot (N_1 + 1) \cdot \dots \cdot (N_d + 1)$; hinzu kommt für jedes der $(N_1 + 1) \cdot \dots \cdot (N_d + 1)$ Feldelemente die Länge des jeweiligen Strings. Veranschaulicht werden soll dies durch eine Beispielrechnung für den schon mehrfach erwähnten Fall eines Arrays $X_g(I)$ (I =Laufindex für die Fälle, N =Anzahl aller Fälle), dessen Elemente Strings mit einer Länge $SL=71$ sind:

$$SP_g = 5 + 2 \cdot 1 + 3 \cdot N + 71 \cdot N = 7 + 74 \cdot N. \quad 1)$$

Ein Vergleich mit der Berechnung auf S. 31 zeigt, daß durch die Verwendung eines Stringarrays pro Fall 266 byte eingespart werden.

-
- 1) Auch die Speicherung in einem zweidimensionalen Integerarray, wie sie im genannten Beispiel möglich wäre, erfordert mehr Platz. Der Vollständigkeit wegen sei auch diese Berechnung hier, allerdings ohne Erklärung der Formel, aufgeführt. Nach der Formel $SP_g = 5 + 2 \cdot d + 2 \cdot (N_1 + 1) \cdot \dots \cdot (N_d + 1)$ ergibt sich: $SP_g = 5 + 2 \cdot 2 + 2 \cdot N \cdot 68 = 9 + 136 \cdot N$.

Die Daten müssen also über INPUT-Statements als String eingelesen werden,¹⁾ können dann geprüft werden und nur die geprüften Strings werden auf Kassette gespeichert. Dieses Vorgehen erfordert, wie aus der FORTRAN-Programmierung und SPSS-Nutzung - dessen Programme in FORTRAN geschrieben sind - bekannt ist, eine Formatangabe, damit der Rechner die Variablen im String auch wiederfindet. Zusätzlich muß also zu Anfang der Dateneingabe, als erster String der Datei, ein Formatstring eingegeben werden, der für jede Variable in der Reihenfolge ihrer Eingabe, die Anzahl der Zeichen enthält, die sie benötigt. Dasselbe kann erfolgen für missing values und Minima und Maxima der Wertebereiche. Falls für eine Variable nicht für alle Fälle Werte vorliegen, so wird diesen Fällen der missing value code der Variablen zugewiesen. Die Fälle mit diesem Code werden in der Regel aus allen weiteren Auswertungen dieser Variablen ausgeschlossen.

Der Formatstring sollte nur einstellige Zahlen enthalten, um das "Aufbrechen" mit Hilfe der Stringfunktionen etwas einfacher zu gestalten. Das bedeutet eine Beschränkung der Variablenwerte auf höchstens 9 Stellen, was für die Praxis ausreichend erscheint.

Das Format der Variablen 10 in der Variablenliste findet man im Programm durch die einfache Befehlsfolge (F% sei der Formatstring, F das

1) Probleme können auftreten bei Strings, die länger als 80 Zeichen sind, da manche Rechner längere Eingaben nicht annehmen bzw. die ersten 80 Zeichen vergessen, obwohl Strings bis zu 256 Zeichen enthalten können. Dann muß der Datenstring zerlegt werden für die Eingabe und nach Eingabe aller Teilstrings wieder zusammengesetzt werden mit "+".

Format der untersuchten Variablen):

$H\$_{SF} = \text{MID}\$(F\$, 1\emptyset, 1): F = \text{VAL}(H\$_{SF})$.

Den Beginn einer Variablen im Datenstring eines Falles I findet man durch Aufsummieren der Formate aller vorhergehenden Variablen (SF) und die Addition von 1, d. h. den Wert der Variablen $1\emptyset$ wie folgt:

$X = \text{VAL}(\text{MID}\$(X\$(I), SF+1, F))$.

Dieses Vorgehen erfordert etwas mehr Programmieraufwand, insbesondere sollte beim Aufruf von Prozeduren darauf geachtet werden, daß die Variablen so weit wie möglich in der Reihenfolge der ursprünglichen Variablenliste ausgewertet werden. Dies ist natürlich nicht sehr benutzerfreundlich, auch dies ein Grund mehr, auf das Verfahren von SCHNELL (1982) zurückzugreifen, da Benutzerfreundlichkeit ein wesentlicher Gesichtspunkt bei der vorliegenden Arbeit ist.

Eine Verbesserung ist auch zu erreichen, wenn man sich entschließt, zu Beginn aller Rechenprozeduren nach Einlesen des Formatstrings und der Datenstrings zunächst alle Formate einem Integerarray $F\$(J)$ (J =Laufindex für die Variable) zuzuweisen und für alle Variablen die Summe aller Formate der vorhergehenden Variablen in ein Integerarray $SF\$(J)$ zu schreiben. Damit stehen diese Zahlen für alle weiteren Prozeduren zur Verfügung und die Behandlung der Variablen kann in beliebiger Reihenfolge geschehen.¹⁾ Der Platz-

1) Der Formatstring $F\$_{SF}$ wird nun nicht mehr benötigt und kann im weiteren Verlauf des Programms überschrieben werden.

bedarf erhöht sich um zweimal $7+2 \cdot M$ (wenn M die Anzahl der Variablen ist), d. h. im Beispiel nur um insgesamt 286 byte, was aber unter Umständen im weiteren Programm eingespart werden kann, da SF nicht mehr berechnet werden muß.

Die Speicherung numerischer Daten als Strings ist sicher nicht die eleganteste Lösung, das Speicherplatzproblem zu lösen, weshalb in der vorliegenden Arbeit ja auch darauf verzichtet wird, doch erscheint sie für viele Anwender, die mit Geräten schlecht ausgerüstet sind, als brauchbare Möglichkeit, das "Problem des knappen Speichers" zu entschärfen. Erst wenn die Länge der Datenstrings fünfmal so groß ist wie die Anzahl der Variablen, also jede Variable im Schnitt 5 Stellen beansprucht, benötigt die Stringspeicherung etwas mehr Platz als die Methode der traditionellen zweidimensionalen Datenmatrix. Dies wird in der Praxis der empirischen Sozialforschung höchst selten einmal der Fall sein.

Sind alle Variablen zweiseitig, sollte auf ein Integerarray zurückgegriffen werden.

Es muß berücksichtigt werden, daß die Auswertungsprogramme für die eine oder andere Speicher-methode nicht die gleichen sein können, also müssen die im System REP-TEST enthaltenen Unterprogramme zur Beurteilung statistischer Daten zumindest verändert werden, wenn die String-speicherung als Permanentspeichermethode gewählt wird. Dies ist nicht sehr schwierig, erfordert aber etwas Zeitaufwand.

2.3 Dateneingabe

Das in 2.2 beschriebene Verfahren der Datenspeicherung hat natürlich Auswirkungen auf die Dateneingabe. Es arbeitet mit Dateien auf Variablenbasis, d. h. die Daten sollten ebenfalls variablenweise (also jeweils eine Spalte der Datenmatrix) eingegeben werden. Dieses ungewohnte Vorgehen erfordert vom Benutzer eine kurze Zeit der Umstellung, es hat aber den Vorteil, daß auf ein Transpositionsprogramm verzichtet werden kann. Bei interaktiver Dateneingabe ist es sinnvoll, auch die Prüfung der Daten auf Plausibilität unmittelbar nach der Eingabe vorzunehmen. Dadurch entfällt die Notwendigkeit eines Fehlersuchprogramms, das nach Abschluß der Dateneingabe einen Teil der fehlerhaften Eingaben aussortiert. Die Fehler-suche ist ebenfalls einfacher bei variablenwei-ser Eingabe, da sich missing values und Wertebereich nach Variablen verändern können. Es ist bekannt, daß ein Großteil der SPSS-Benutzer erst nach den ersten Druckerprotokollen anhand der Häufigkeitstabellen Fehler in den Daten entdeckt, wenn nämlich Häufigkeiten von Werten auftauchen, die außerhalb des zulässigen Bereichs einer Variablen liegen. Die nachträgliche Suche nach Fehlern kostet dann besonders bei großen Datenbeständen viel Zeit.

Es erscheint daher sinnvoll, für die Dateneingabe ein wenig mehr Zeit aufzuwenden, dafür aber dann auch nur geprüfte Daten zu speichern. Dies leisten die beiden Programme FILE INPUT und DATA INPUT.

Das Programm FILE INPUT definiert die Datei der sequentiellen Files der Variablenwerte und kreiert ein File mit dem Namen VARCATALOG, das auf die als DATA DISK bezeichnete Datendiskette geschrieben wird. Hier sind alle Informationen über die Datei enthalten in der Reihenfolge:

1. Dateiname F%
2. Zahl der Variablen NV%
3. Zahl der Fälle NF%
4. Missing value MW
5. Statuspointer für die Variablen IS%
6. Statuspointer für die Fälle IN%
7. Variablennamen nach der Reihenfolge
der Eingabe

Es ist zu beachten, daß nur mit einem missing value für alle Variablen gearbeitet wird.

Der Dateiname wird genau wie die Variablennamen auf seine Länge geprüft, die sechs Zeichen nicht überschreiten darf. Andernfalls wird erneute Eingabe verlangt. Eine neue Eingabe wird ebenfalls angefordert, falls das erste Zeichen des Namens kein Buchstabe ist.

Die Zahl der Variablen muß größer als 1 und kleiner als 81 sein, die Zahl der Fälle größer als 1 und kleiner als 201, ansonsten muß die Eingabe der Werte für NV% und NF% wiederholt werden.

An dieser Stelle werden alle Parameter noch einmal angezeigt und können alle neu eingegeben werden.

Nach Eingabe aller Variablennamen werden sie noch einmal aufgelistet und können durch die LABEL CORRECTION-Subroutine verändert werden. Falls Veränderungen nötig sind, werden erneut die Länge und das erste Zeichen geprüft.

Führen Bedienungsfehler zu "FATAL ERROR"-Meldungen, muß das Programm erneut gestartet werden; die eingegebenen Parameter sind verloren. Das Programm hat einen Umfang von 5 Kbyte, man benötigt 7 Sekunden, um es zu laden.

Nach dem Schreiben des Files VARCATALOG auf die Datendiskette erfolgt der Rücksprung zum Hauptprogramm.

Nun kann der Aufruf des DATA INPUT-Programms erfolgen zur Eingabe der Rohdaten. Der Programmablauf soll hier nur ganz kurz beschrieben werden.¹⁾

Es wird zunächst geprüft, ob entsprechend der Dateidefinition im VARCATALOG noch weitere Variablen einzugeben sind, d. h. die Dateneingabe kann nach einem Teil der Variablen abgebrochen werden und zu einem späteren Zeitpunkt fortgesetzt werden; es kann auch nach einer Anzahl von Fällen für eine Variable unterbrochen werden, da die Prüfung des Statuspointer IN% bei Beginn der Dateneingabe erfolgt. Ist IN% ungleich eins, werden alle für diese Variable bereits gespeicherten Daten eingelesen und man kann an derselben Stelle fortfahren. Für jede Variable kann zu Beginn ein "Range-check" gewählt werden. Es werden dann

1) Ausführliche Erläuterungen finden sich in der Arbeit von SCHNELL (1982), S. 50-57, ebenso zu FILE INPUT.

Minimum und Maximum angefordert und alle eingegebenen Werte daraufhin geprüft, ob sie innerhalb des zulässigen Bereichs liegen. Vorher erfolgt eine Prüfung, ob ein EXTRA-IGNORED-Fehler aufgetreten ist, d. h. ob statt des Dezimalpunkts ein Komma eingegeben wurde. Da für fehlende Werte der Buchstabe "M" einzugeben ist, erfolgt das Einlesen der Daten in die Stringvariable A\$, ein Vergleich von VAL(A\$) mit STR\$(VAL(A\$)) ermöglicht es z. B., führende Nullen bei der Eingabe als fehlerhaft zurückzuweisen. "M" wurde bereits vorher durch den im VARCATALOG genannten missing value ersetzt. Falls der nun folgende Range-check günstig verlaufen ist, wird der Wert von A\$ den Arrays A(I) und IO(I) zugewiesen.

Nach jeweils 10 Eingaben können diese letzten Eingaben noch editiert werden, erkennbar an dem "MODE"-Zeiger in der rechten oberen Bildschirm-ecke, der von INPUT DATA zu CONTROL DATA wechselt; soll editiert werden, erscheint EDIT DATA.

Sind alle Fälle für eine Variable eingegeben, werden diese Werte auf der Diskette gespeichert, und nach der Abfrage "CONTINUE INPUT?" wird die nächste Variable eingegeben bzw. zum Hauptprogramm zurückgesprungen.

Sollen mehr Fälle oder Variablen aufgenommen werden, als ursprünglich vorgesehen war, muß über das Programm DATA BASE SYSTEM zunächst der VARCATALOG geändert werden, anschließend kann mit DATA INPUT wieder eingegeben werden.

Das Programm beansprucht ca. 9 Kbyte, das Laden benötigt 9 Sekunden.

2.4 Datenkorrektur

Das System DATAN II arbeitet mit einem Data Base Mangement System (DBMS), das umfangreiche Datenmodifikationen ermöglicht. Seine vielfältigen Leistungen detailliert aufzuzählen, fehlt hier der Platz.¹⁾ Das Programm hat einen Umfang von knapp 15 K, es benötigt insgesamt 33 K. Damit ist der Arbeitsspeicher des APPLE fast vollständig ausgenutzt. Anhand des Menus dieses Programms soll die Vielseitigkeit kurz aufgezeigt werden:

DATAN II DATA BASE MANAGEMENT SYSTEM	
DATA INPUT	DATA EDITING
FILE INPUT.....1	LIST CASES.....3
DATA INPUT.....2	EDIT DATA.....4
FILE MANAGEMENT	DATA MODIFICATION
ADD CASES.....5	COMPUTE.....10
ADD VARIABLES....6	COUNT.....11
DELETE VARIABLE..7	RECODE.....12
DELETE FILE.....8	IF.....13
COPY FILE.....9	Z-TRANSFORMATION.14
CLEAR MEMORY.....15	MAIN MENU.....17
FILE CATALOG.....16	WHICH OPTION.....

1) Die Programmbeschreibung bei SCHNELL (1982) benötigt 25 Seiten.

Da es auch für Benutzer von REP-TEST nützliche Prozeduren enthält, wird es als ein Programmblock in das System mit aufgenommen. Insbesondere die Optionen 1 bis 9 sind interessant.

Wie am Ende des Abschnitts 2.3 kurz angesprochen, ist es manchmal nötig, die Variablenliste zu erweitern oder die Fallzahl zu vergrößern. Dies leisten die Prozeduren ADD CASES (5) und ADD VARIABLES (6); LIST CASES (3) ist sehr hilfreich beim Suchen von Fehlern und EDIT DATA (4) unbedingt nötig zur Korrektur von Daten.

FILE INPUT (1) und DATA INPUT (2) sind schon näher erläutert worden.

Die Optionen 10 - 14 sind die dem SPSS-Benutzer bekannten:

COMPUTE berechnet neue Variablen aus den gegebenen.

COUNT zählt die Häufigkeit von Werten über alle Variablen aus (pro Fall) statt die Häufigkeit über alle Fälle (pro Variable).

RECODE ermöglicht Umcodierungen der Variablenwerte.

IF ermöglicht das Schaffen neuer Variablen, deren Werte von der Erfüllung einer Bedingung (oder mehrerer) abhängen.

Z-TRANSFORMATION leistet die bekannte Standardisierung einer Variablen zu einer neuen Variablen mit Mittelwert gleich 0 und Standardabweichung gleich 1.

CLEAR MEMORY löscht den Inhalt der CPU mit Ausnahme des Programms DATA BASE SYSTEM, wenn nacheinander verschiedene Dateien zu bearbeiten sind. Sonst müßte das System neu gestartet werden.

LIST CASES ermöglicht die Darstellung von höchstens 15 Variablenwerten eines Falles am Bildschirm.

EDIT DATA ermöglicht Änderungen der Daten nach der Abspeicherung auf die Diskette. Für die zu editierende Variable kann der entsprechende Fall durch die Fall-Nummer oder einen Rangecheck ermittelt werden. Nach etwaigen Korrekturen eines oder mehrerer Fälle erfolgen aber keine Eingabeprüfungen mehr wie im Programm DATA INPUT (d. h. Kommaeingabe statt Punkt oder Buchstaben statt Zahlen), deshalb darf in diesem Unterprogramm "M" für den missing value nicht verwandt werden.

Die Benutzung dieser Verfahren, insbesondere von COMPUTE und IF, setzt allerdings beim Benutzer höhere Vorkenntnisse voraus, was die Syntax von BASIC-Formeln und -Funktionen angeht, in etwa gleichem Maße wie die SPSS-Nutzung Kenntnisse der besonderen Steuersprache verlangt.

ADD CASES und ADD VARIABLES erlauben eine Erweiterung der existierenden Datei um zusätzliche Fälle bzw. Variablen. Im Anschluß an diese beiden Unterprogramme muß das DATA INPUT-Programm aufgerufen werden, um die noch fehlenden Werte einzugeben. Beide Programme prüfen, ob die aus dem Programm FILE INPUT bekannten Beschränkungen ($2 \leq NV \leq 80$, $2 \leq NF \leq 200$) durch die Erweiterungen überschritten werden. Das Programm ADD VARIABLES enthält ebenfalls - wie FILE INPUT - eine LABEL CORRECTION-Routine.

DELETE VARIABLE löscht den kompletten Datenfile einer angegebenen Variablen auf der Diskette. Da

diese Werte dann verloren sind für weitere Auswertungen, erfolgt noch eine Sicherheitsabfrage, ob wirklich gelöscht werden soll. Hierbei ändert sich der Variablenkatalog, da die jeweils letzte Variable des Katalogs an die Stelle der gelöschten tritt. Damit wird die Reihenfolge der Variablen im Katalog verändert.

DELETE FILE löscht die gesamte Datei aller Variablenfiles auf der Diskette, auch den Variablenkatalog.

COPY FILE kopiert den Inhalt einer Diskette auf eine andere. Hierzu wurde das in ASSEMBLER geschriebene Kopierprogramm der APPLE SOFT MASTER DISKETTE um eine Rückkehrroutine zum DATA BASE SYSTEM erweitert. Da nur ein Laufwerk zur Verfügung stand, muß die Zieldiskette als Slot 6 mit Drive 1 angegeben werden.

Diese kurzen Informationen zum DATA BASE SYSTEM von DATAN II genügen, um den Benutzer von REPTEST in die Lage zu versetzen, die zur Datenkorrektur notwendigen oder wünschenswerten Prozeduren zu nutzen. Für anspruchsvollere Anwender wurden auch die übrigen Verfahren des DBMS übernommen.

Kurz vor Abschluß der vorliegenden Arbeit wurde von SCHNELL eine gekürzte und verbesserte Version des Systems DATAN, genannt DATAN II, fertiggestellt, die aber leider nicht mehr für REPTEST genutzt werden konnte.

Kap. 3 Methoden der induktiven Statistik

3.1 Aufgaben der induktiven Statistik

Die vorliegende Arbeit beschränkt sich auf ein Teilgebiet der Statistik, nämlich die induktive. Die Deskription von Datenbeständen bleibt ausgespart, weil zum Teil schon gute Programme für diese Aufgaben für Kleinrechner vorliegen. Auch wenn diese nicht immer den gewünschten Grad an Benutzerfreundlichkeit aufweisen, fallen die in der Einleitung geschilderten Probleme der Auswahl der geeigneten Verfahren im Bereich der deskriptiven Statistik nicht so schwer ins Gewicht wie bei der induktiven Statistik. Die Beschreibung von Datenbeständen in Häufigkeitstabellen, mit univariaten und bivariaten Kennzahlen ist auch für den weniger geübten Anwender eher zu durchdringen als das kompliziertere Gebiet der schließenden Statistik.

3.1.1 Signifikanztests

Die beiden Verfahrensgruppen der Inferenzstatistik "Testen" und "Schätzen" sind zusammen immer noch zu umfangreich, um im Rahmen dieser Arbeit behandelt zu werden. Deshalb wurde ausschließlich der Bereich der Testtheorie ausgewählt, der zudem bei den bisherigen Programmveröffentlichungen aus der Statistik in Quantität und Qualität weit hinter den Verfahren der deskriptiven Statistik

zurückgeblieben ist. Und das obwohl die Signifikanztests¹⁾ sich immer weitere Anwendungsgebiete erobert haben. Dies hat einen Grund auch in der Skepsis gegenüber den Ergebnissen der deskriptiven Statistik; diese führt doch nur zu vorläufigen Aussagen, zu Hypothesen über den empirischen Sachverhalt. Forschungen aus den Bereichen Meßtheorie und Instrumententheorie haben erheblich dazu beigetragen, daß drei wesentliche Fehlerquellen als solche erkannt wurden, deren Auswirkungen die tatsächlichen Verhältnisse verschleiern können. Diese sind:

1. Fehler des Beobachters,
2. Fehler des Beobachtungsinstruments und
3. Fehler bei der Auswahl der beobachteten Fälle.

Das Wissen um die möglichen Fehler führte natürlich zu einer kritischeren Betrachtung von Untersuchungsergebnissen und zu der Frage, ob beobachtete Effekte auf die Fehlerquellen oder auf empirische Sachverhalte zurückzuführen sind.

Hier nun ist der Ansatzpunkt für den Einsatz von Testverfahren. Sie stellen ein Hilfsmittel dar, zur Entscheidung der Frage, ob ein konkretes Meßergebnis "durch Zufall", d. h. durch unkontrollierbare Fehlereinflüsse, zustande gekommen ist oder durch systematische Einflüsse. Dabei hat man wiederum Möglichkeiten der Fehlentscheidung, nämlich die bekannten Fehler erster und zweiter Art, d. h. ein Ergebnis als Folge von Fehlerein-

1) Eine interessante kritische Diskussion der Signifikanztestverfahren findet sich bei WITTE, E. H.: Signifikanztest und statistische Inferenz, Stuttgart 1980

flüssen ansehen, obwohl es eine systematische Ursache hat (β -Fehler), bzw. ein Ergebnis als Folge systematischen Einflusses ansehen, obwohl es durch Fehlerquellen verursacht wird (α -Fehler).¹⁾ Das Vorgehen bei Anwendung von Tests sieht im allgemeinen so aus:

Aufgrund von theoretischen Überlegungen kann in einer konkreten Meßsituation ein bestimmtes Ergebnis erwartet werden, wenn das theoretische Modell der empirischen Situation angemessen ist. Dieses erwartete Ergebnis wird als Hypothese formuliert.²⁾ "Kleinere" Abweichungen des tatsächlichen Ergebnisses von der Hypothese werden als zufällig angesehen, "größere" hingegen als Grund für die Ablehnung der Hypothese, d. h. als Grund für die Vermutung, daß der theoretisch postulierte Einfluß nicht vorliegt. Dem liegt der Gedanke zugrunde, daß bei zutreffender Hypothese, d. h. für den Fall, daß die vermuteten, theoretisch abgeleiteten Einflüsse tatsächlich auftreten, die Wahrscheinlichkeit einer "größeren" Abweichung von dieser Hypothese so gering ist, daß das Auftreten eines solchen Ergebnisses als Anhaltspunkt - nicht Beweis - für das Nicht-Zutreffen der Hypothese gewertet werden muß; die Abweichung wird dann als signifikant bezeichnet. Bisher ist noch offen geblieben, was größere oder kleinere Ab-

-
- 1) Die Bezeichnungen α - und β -Fehler sind entstanden, da die Wahrscheinlichkeiten für diese Fehler in der Literatur fast durchgängig mit α bzw. β bezeichnet wurden.
 - 2) Hier haben sich 2 Begriffspaare entwickelt: Hypothese/Alternative bzw. Nullhypothese/Alternativhypothese als Gegensatzpaare. Diese Arbeit benutzt das erste Paar in Anlehnung an die traditionellere Literatur der mathematischen Statistik.

weichungen sind. Hier ist nun - wie bekannt ist - ein Aspekt der Willkür den Signifikanztestverfahren eigen. Die Willkür liegt in der Wahl des Signifikanzniveaus, das ist die Wahrscheinlichkeit α des Fehlers erster Art. Der sogenannte Rückweisungsbereich eines Tests wird so festgelegt, daß unter den theoretisch abgeleiteten Voraussetzungen die Wahrscheinlichkeit dafür, daß ein Ergebnis aus diesem Bereich auftritt, gerade gleich α ist.¹⁾ Dabei wird α möglichst klein gewählt, d. h. 5 %, 2 % oder 1 %. Je nach Wahl von α kann die Testentscheidung unterschiedlich lauten: Ablehnung oder Beibehaltung. Dies soll im Auge behalten werden, aber nicht das Verdienst von Signifikanztests schmälern, zumal in vielen Fällen der Willkür Schranken gesetzt sind durch die ökonomischen oder sozialen Konsequenzen einer Fehlentscheidung. Auch die Asymmetrie des Entscheidungsprozesses soll erwähnt werden. Nur die Rückweisung einer Hypothese bei niedrigem Signifikanzniveau kann als Erkenntnisfortschritt angesehen werden, weshalb Hypothese und Alternative im Grunde oft als vertauscht angesehen werden können. Ein Beispiel soll verdeutlichen, was gemeint ist: Vermutet man einen Zusammenhang zwischen zwei Merkmalen einer Population (dies ist also die eigentliche Hypothese), so wird als Testhypothese die Unabhängigkeit formuliert. Starke Abweichungen des empirischen Befundes von dem bei Unabhängigkeit zu erwartenden Ergebnis haben bei kleinem α eine so geringe Wahrscheinlichkeit, daß die Ablehnung der Hypothese "Unabhängigkeit" fast mit Sicher-

1) Bei diskreten Verteilungen kann α oft nicht erreicht werden (oder müßte überschritten werden).

richtig ist.¹⁾ Hingegen sagt eine Beibehaltung²⁾ der formulierten Hypothese beinahe nichts aus, da immer noch der Fehler zweiter Art aufgetreten sein kann, d. h. eine falsche Hypothese wurde beibehalten.

Zusammengefaßt: Bei Ablehnung der Hypothese "Unabhängigkeit" ist die eigentliche Vermutung, die auf Zusammenhang lautete, stark gestützt worden und kann je nach Repräsentativität und Umfang der Stichprobe, bei geringem Signifikanzniveau des Tests, als fast sicher angesehen werden. Im anderen Fall müssen weitere Untersuchungen folgen.

Signifikanztestverfahren können also keine Sicherheit bieten, aber doch zu einem erheblichen Maß an Informationsgewinn führen. Sie sollten deshalb auch Anwendern zugänglich sein, die sich bisher nur auf Tabellen, Mittelwerte und Streuungsmaße beschränkt haben. Sie haben ihre Berechtigung bei der Absicherung von deskriptiv-statistischen Befunden; Fehlereinflüsse können unter Umständen mit größerer Sicherheit angenommen werden. Also müssen Statistikprogrammpakete auch inferenzstatistische Methoden umfassen, auch solche Systeme, die für Kleinrechner konzipiert wurden. Daß im Hinblick auf den Anwenderkreis von Mikrocomputern noch weitere Aspekte berücksichtigt werden müssen, kann nicht bedeuten, daß auf solche Verfahren verzichtet werden sollte.

-
- 1) Hier ist natürlich nicht der wahrscheinlichkeitstheoretische Begriff "fast sicher" gemeint, der eine Wahrscheinlichkeit von 1 voraussetzt bei stetiger Verteilung.
 - 2) Dieser Begriff wird dem der Annahme einer Hypothese vorgezogen, obwohl er sprachlich unschön ist. Annahme hingegen trifft nicht den Sachverhalt.

3.1.2 Parametrische und nicht -parametrische Tests

Die Einteilung der Testverfahren in parametrische und nicht-parametrische ist im Grunde etwas irreführend, da z. B. der Median-Test, der zu den nicht-parametrischen Verfahren zählt, durchaus einen Parameter testet. Der Unterschied läßt sich vielmehr folgendermaßen darstellen:

Parametrische Testverfahren setzen eine bestimmte Annahme über die stetige Verteilung einer Variablen in der Grundgesamtheit voraus, zumindest über die Zugehörigkeit zu einer "Familie" von Verteilungen. Getestet werden Hypothesen über charakteristische Parameter, die die Verteilung bestimmen, z. B. Erwartungswert oder Varianz.

Der Vorteil dieser Tests liegt in der hohen Güte für den Fall, daß die Verteilungsannahme exakt erfüllt ist. Die Gütefunktion $1-\beta(\theta)$ ist ein wichtiges Kriterium für die Auswahl eines Testverfahrens, wenn mehrere Tests möglich sind. Dabei ist $\beta(\theta)$ die Wahrscheinlichkeit dafür, eine falsche Hypothese beizubehalten; diese Wahrscheinlichkeit ist natürlich abhängig davon, welcher Wert θ der Alternative der zutreffende ist. Die Güte eines Tests, $1-\beta(\theta)$, wird somit angegeben durch die Wahrscheinlichkeit, eine falsche Hypothese auch tatsächlich zurückzuweisen.

Eine weitere Voraussetzung für die Anwendung "verteilungsgebundener" Verfahren muß noch genannt werden: Sie setzen metrisches Skalenniveau voraus und sind deshalb für viele Anwendungen innerhalb der empirischen Sozialforschung nicht geeignet.

Aus diesem Grund soll in dieser Arbeit vorrangig die zweite Gruppe von Tests behandelt werden, die nicht-parametrischen oder verteilungsfreien Verfahren. Sie können schon bei ordinalem Meßniveau eingesetzt werden, viele auch bei nominalem. Die Anwendung dieser Verfahren setzt keine Annahme über die Verteilung in der Grundgesamtheit voraus, zumindest keine detaillierte Annahme. Manchmal wird mindestens die Stetigkeit der Verteilung verlangt oder Gleichheit der Verteilung in zwei verschiedenen Grundgesamtheiten.

Die insgesamt wesentlich schwächeren Voraussetzungen machen diese Verfahren für die Sozialwissenschaften so interessant. Auch für den Fall metrischen Skalenniveaus sind nicht-parametrische Tests dann einzusetzen, wenn eine Verteilungsannahme nicht gemacht werden kann, allenfalls eine vage Vermutung zur Grundlage des weiteren Vorgehens gemacht werden könnte. Es ist zu sagen, daß verteilungsgebundene Tests generell eine höhere Güte aufweisen als verteilungsfreie, wenn die Verteilung der Grundgesamtheit bekannt ist - bei gleichem Stichprobenumfang. Eine Verbesserung der Güte kann allerdings immer durch Erhöhung des Stichprobenumfangs erreicht werden. Aber Überlegungen über möglicherweise geringere Güte eines Verfahrens sind völlig unangebracht, wenn das vorliegende Skalenniveau zu niedrig ist für parametrische Tests - wie meistens im Bereich der Sozialwissenschaften. Die Tatsache, daß die Auswahl der geeigneten statistischen Methode abhängig ist vom Meßniveau, ist im übrigen auch Anwendern deskriptiver Methoden bekannt. Die Beschreibung der Testverfahren

ren in den folgenden Abschnitten enthält im wesentlichen eine knappe Beschreibung der Arbeitsweise der Testverfahren und die wichtigsten Voraussetzungen für ihre Anwendung, die Gliederung orientiert sich an Problemtypen der Praxis.¹⁾

In das System REP-TEST wurden ausschließlich verteilungsfreie Verfahren aufgenommen, da sie - besonders von Sozialwissenschaftlern - wesentlich häufiger eingesetzt werden können wegen der genannten Gründe. Das System bietet damit in etwa den Standard der Prozedur NPAR TESTS des Programmpaketes SPSS, mit Ausnahme des wenig gebräuchlichen MOSES-TESTS, eines Anpassungstests bei ordinalem Meßniveau. REP-TEST enthält zudem noch die Prüfung von Zusammenhängen zwischen Variablen, die in NPAR TESTS nicht enthalten sind. Damit ist das Angebot so umfangreich, daß die Aufnahme weiterer Verfahren, z.B. auch parametrischer, bei der zur Verfügung stehenden Gerätekonfiguration unmöglich erschien. Das Programm "Tests mit Hilfe", das weniger erfahrenen Anwendern Hilfestellung bieten soll, enthält nicht alle der im Programm "Testverfahren" zur Verfügung stehenden Verfahren, es wurde eine Vorauswahl getroffen, damit eine eindeutige Zuordnung der Tests zu bestimmten Problemsituationen möglich wurde. Mit einer klaren Strukturierung ist diesem bestimmten Kreis von Benutzern eher gedient als mit einem möglichst reichhaltigen Angebot.

1) Es konnte natürlich nur ein kleiner Teil der nicht-parametrischen Verfahren ausgewählt werden, deren Entwicklung nach grundlegenden Arbeiten, z. B. von FRIEDMAN, SMIRNOW, KENDALL, WALD, in den 30er und 40er Jahren so rasch fortschritt, daß SAVAGE in seiner "Bibliography of Nonparametric Statistics" schon etwa 3000 Titel nannte.

3.2 Prüfung einer empirischen Verteilung auf Abweichung von einer definierten Verteilungsform

In diesem Abschnitt sind Tests zusammengefaßt, die für den Einstichprobenfall eine gegebene empirische Häufigkeitsverteilung auf Abweichung von einer "theoretischen" Grundgesamtheitsverteilung prüfen. Dies läuft im Falle einer diskreten Verteilung auf die Prüfung der Anteilswerte hinaus, d. h. auf die Prüfung der Frage, ob die beobachteten Anteilswerte der Stichprobe von gegebenen Anteilswerten der Grundgesamtheit abweichen, und zwar signifikant abweichen. Als Hypothese wird die Gleichheit der beiden Werte angenommen bzw. die Gleichheit der jeweils einander entsprechenden Anteilswerte im polytomen Fall. Bei einer stetigen Verteilung, also mindestens intervallskalierten Daten, werden die theoretische und die empirische Verteilungsfunktion miteinander verglichen. Wiederum setzt die Hypothese die Gleichheit der Verteilungsfunktionen voraus, wie bei allen Anpassungstests.

Binomialtest

Dieser Anteilswerttest ist einzusetzen bei dichotomen Variablen von beliebigem Skalenniveau.

Die Verteilungsfunktion einer solchen Variablen ist durch die Angabe des Anteilswertes π einer der beiden Ausprägungen vollständig beschrieben, d. h. die allgemein formulierte Hypothese eines Anpassungstests - $F(x) = F_0(x)$ - kann umgewandelt werden zu $H_0: \pi = \pi_0$.

Der Stichprobenanteilswert P führt zu einer Prüfgröße mit angebbarer Verteilung.¹⁾ Bei einem Stichprobenumfang von n folgt die Prüfgröße $n \cdot P$ einer $B(n; \pi_0)$ -Verteilung, falls die Hypothese zutreffend ist. Der Test kann ein- oder zweiseitig sein, d. h. die Alternative kann lauten:
 $H_a : \pi \neq \pi_0$ bzw. $H_a : \pi > \pi_0$ oder $H_a : \pi < \pi_0$.
 Die Bestimmung der Grenzen des Annahmebereichs, der Rückweisungspunkte, ist für eine Binomialverteilung aufwendig, für kleine Stichprobenumfänge n allerdings unumgänglich. Dabei ist es durchaus möglich, daß das Signifikanzniveau nicht erreicht wird - wie oft bei diskreten Verteilungen. Bei großem n kann auf die Normalverteilung ausgewichen werden. Die Begründung liefert der Grenzwertsatz von DE MOIVRE/LAPLACE:²⁾

Mit wachsendem n nähert sich die Binomialverteilung $B(n; \pi)$ einer Normalverteilung an mit den Parametern $\mu = n \cdot \pi$ und $\sigma^2 = n\pi(1-\pi)$, d. h. für die

-
- 1) Es wird angenommen, daß die Voraussetzungen des BERNOULLI-MODELLS erfüllt sind, d. h. es dürfen nur 2 einander ausschließende Ausprägungen (1/0, A/ $\bar{A}$ o. ä.) auftreten, für jedes Element der Grundgesamtheit müssen die Wahrscheinlichkeiten für 1 (oder A) gleich sein (Konstanz von π) und das Ergebnis für ein Element ist unabhängig von dem der übrigen (Modell mit Zurücklegen). Dieser letzte Punkt ist im allgemeinen nicht erfüllt, doch bei ausreichend großen Grundgesamtheiten für die Praxis vernachlässigbar.
 - 2) Diese Aussagen über Verteilungskonvergenz gehen auf das zentrale Grenzwerttheorem zurück. Siehe dazu z. B. WITTING, H. / NOELLE, G.: Angewandte Mathematische Statistik, Stuttgart 1970; für die mathematischen Grundlagen dieser Approximation z. B. BAUER, H.: Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie, Berlin 1968

Wahrscheinlichkeitsfunktion $P(x)$ der $B(n, \pi)$ -Verteilung gilt:

$$P(x) \sim f(x) = \frac{1}{\sqrt{n \cdot p(1-p)}} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\frac{x-np}{\sqrt{np(1-p)}} \right)^2}$$

(In dieser Formel ist π die Kreiskonstante, p der Anteilswert.)

Für die Bestimmung der Rückweisungspunkte mit Hilfe der Verteilungsfunktion der Normalverteilung sollte eine "Stetigkeitskorrektur" von 0,5 berücksichtigt werden, d. h. die Punkte p_{r_i} , $i = 1, 2$, ergeben sich zu:

$p_{r_i} = \sqrt{n \pi_0 (1 - \pi_0)} \cdot k_{r_i} + n \cdot \pi_0 \pm 0,5$, wobei k_{r_i} die Rückweisungspunkte der Standardnormalverteilung bei gleichem Signifikanzniveau sind.

Es bleibt die Frage zu klären, wann der Stichprobenumfang n als ausreichend groß für die Verwendung der Normalverteilung angesehen wird. Wie meist in solchen Fällen gibt es eine Fülle von Faustregeln für die Größe von n ; manche geben nur ein Minimum für n an, andere machen dies noch abhängig von der Größe von π . Hier wird die auch von SPSS benutzte Regel verwandt, daß ab $n > 25$ die exakte Binomialverteilung durch die Normalverteilung approximiert werden kann. Dieser groben Regel wurde der Vorzug gegeben gegenüber der von TIEDE/VOSS¹⁾ genannten Regel, wonach $n > \frac{9}{p(1-p)}$ als Kriterium für eine ausreichend gute Näherung anzusehen ist. Hier waren Konzessionen

1) TIEDE, M./VOSS, W.: Induktive Statistik, Teil 1, Köln 1977

nötig an die Möglichkeiten eines Mikrocomputers, dessen Rechengrenzen bei Binomialkoeffizienten leicht erreicht werden. Durch geschicktes Programmieren wurde erreicht, daß immerhin noch $\binom{50}{24}$ zu bestimmen ist. Da dann aber auch die Rechenzeiten entsprechend lang sind, wurde als Kompromiß $n = 25$ als Grenze gewählt.

Zusammenfassung:

Anzahl der Stichproben	eine
Meßniveau	nominal
Untersuchungsvariable	diskret - dichotom
Hypothese	$\pi = \pi_0$
Alternative	ein- oder zweiseitig
Prüfgröße	$n \cdot P$
Stichprobenverteilung ¹⁾	$B(n, \pi)$ ($\sim N(\pi n; \sqrt{n\pi(1-\pi)})$ f. $n > 25$)

χ^2 -Anpassungstest

Wie bei jedem Anpassungstest lautet auch die Hypothese des χ^2 -Tests $H_0: F(x) = F(x_0)$. Der χ^2 -Test ist anwendbar für diskrete Untersuchungsvariablen, die aber im Gegensatz zum Binominaltest durchaus mehr als zwei Ausprägungen haben können, bei beliebigem Skalenniveau. Die Verteilungsfunktion einer polytomen Variablen ist durch die Angabe der Anteilswerte aller Kategorien eindeutig beschrieben; analog zum Binominaltest formulieren wir also die Hypothese als $H_0: \pi_i = \pi_{0i}, i=1, \dots, r; \sum_i \pi_{0i} = 1$.

1) Gemeint ist natürlich die Stichprobenverteilung der Prüfgröße.

Die r-dimensionale Stichprobenfunktion $(X_1, \dots, X_r)$, die die Anzahl der Elemente in der i-ten Kategorie angibt, folgt einer Multinomialverteilung.¹⁾

Die eindimensionalen Funktionen X_i (Anzahl der Elemente in der i-ten Kategorie) sind asymptotisch (d.h. für $n \rightarrow \infty$) normalverteilt nach einer $N(n\pi_i; \sqrt{n\pi_i})$ -Verteilung, d.h. $(X_i - n\pi_i) / \sqrt{n\pi_i}$ ist standardnormalverteilt.

Die im χ^2 -Test verwendete, von PEARSON²⁾ vorgeschlagene und nach ihm benannte Prüfgröße

$$U = \sum_{i=1}^r \frac{(X_i - n\pi_i)^2}{n\pi_i} \quad \text{ergibt sich somit als}$$

Summe von quadrierten, asymptotisch standardnormalverteilten Variablen. Aus der Herleitung der χ^2 -Verteilung ist bekannt, daß sie die Verteilung einer Summe von quadrierten $N(0,1)$ -verteilten Variablen darstellt. Die Anzahl der Freiheitsgrade, der explizite Parameter der χ^2 -Verteilung, ist gleich der Anzahl der Faktoren in der Summe, wenn alle Variablen unabhängig sind. Wegen der Beschränkung $\sum \pi_i = 1$ bzw. $\sum X_i = n$ hat die asymptotische Verteilung der Pearson'schen Prüfgröße U nicht r, sondern nur r-1 Freiheitsgrade.

1) Gelten ähnliche Voraussetzungen wie im Bernoulli-Modell, erweitert auf den polytomen Fall, so gilt für die Funktion $(X_1, \dots, X_r)$ mit X_i =Anzahl der Stichprobenelemente in der i-ten Kategorie:

$$P(x_1, \dots, x_r) = \frac{n!}{x_1! \dots x_r!} \pi_1^{x_1} \dots \pi_r^{x_r}, \text{ wenn die } \pi_i \text{ die Anteilswerte in der Grundgesamtheit sind.}$$

2) PEARSON, K.: On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. In: Philosophical Magazine Series 5, Vol. 50 (1900), S. 157-172

Die Konstruktion von U geht aus von den Abweichungen der beobachteten Häufigkeiten X_i von denen, die man bei zutreffender Hypothese erwartet, nämlich $n \cdot \pi_i$.¹⁾ Diese werden quadriert, damit sie einander nicht aufheben. Sind die erwarteten Werte groß, so sind größere Abweichungen möglicherweise immer noch als zufällig anzusehen, bei kleinen erwarteten Werten hingegen kann eine Abweichung derselben Größe durchaus signifikant sein. Deshalb wird die Abweichung "relativiert", es wird nur der "prozentuale Fehler" in die Prüfgröße einbezogen.

Es wurde schon festgestellt, daß die Richtung der Abweichung - wegen der Quadrierung - keinen Einfluß hat auf die Prüfgröße, somit auch nicht auf die Testentscheidung. Anders gesagt: Der χ^2 -Test ist immer ein zweiseitiger Test, obwohl nur ein Rückweisungspunkt bestimmt wird, da U immer positiv ist.

Bei der Verwendung des χ^2 -Tests muß berücksichtigt werden, daß U nur asymptotisch χ^2 -verteilt ist. Damit die Näherung ausreichend gut ist, sollte die Faustregel: $n\pi_i \geq 5$ nicht verletzt werden. Gegebenenfalls müssen Kategorien zusammengelegt werden. Ein weiterer Hinweis ist nötig für den Fall, daß die Verteilungshypothese nicht vollständig spezifiziert ist. Dies kann der Fall sein, wenn die Anpassung an eine theoretische Verteilung geprüft werden soll, deren Form man

1) Dies ist ein Grund dafür, daß U oft in der folgenden Notation aufgeführt wird:

$$U = \sum_{i=1}^r \frac{(B_i - e_i)^2}{e_i}, \quad B_i \text{ für beobachtete, } e_i \text{ für erwartete Häufigkeiten.}$$

angeben kann, deren explizite Parameter aber aus der Stichprobe geschätzt werden müssen, damit die π_{oi} der Hypothese exakt bestimmt werden können.¹⁾

Dadurch vermindert sich die Anzahl der Freiheitsgrade um die Anzahl der geschätzten Parameter; das bedeutet, daß die Hypothese eher zurückgewiesen wird, der Annahmehereich ist bei gleichem Signifikanzniveau kleiner. Im Fall eines dichotomen Merkmals liefert der χ^2 -Test dasselbe Ergebnis wie der Binomialtest.

Bei höherem als nominalem Meßniveau kann ein χ^2 -Test ebenfalls sinnvoll sein, wenn keine weitergehende Verteilungsannahme gemacht werden kann; allerdings werden Informationen verschenkt über Größe oder Anordnung der Werte, da nur gruppiertes Material verwendet wird und nur die Häufigkeiten interessieren.

Zusammenfassung:

Anzahl der Stichproben	eine
Meßniveau	nominal
Untersuchungsvariable	diskret - polytom
Hypothese	$\pi_i = \pi_{oi}, i=1, \dots, r;$
Alternative	zweiseitig

Prüfgröße	$U = \sum_{i=1}^r \frac{(X_i - n\pi_i)^2}{n \cdot \pi_i}$
-----------	---

Stichprobenverteilung	$\sim \chi^2(r-1-\delta)$
-----------------------	---------------------------

(Für $n \cdot \pi_i \geq 5$ und δ =Anzahl der geschätzten Parameter).

1) Ein Beispiel ist die Poisson-Verteilung, deren Parameter aus der Stichprobe geschätzt wird.

Kolmogoroff-Smirnow-Anpassungstest

Ein weiterer Test zur Prüfung der Anpassung einer empirischen Verteilung an eine theoretische ist der Kolmogoroff-Smirnow-Test, der für stetige Untersuchungsvariablen verwendet wird. Das bedeutet, daß strenggenommen metrisch skalierte Daten vorliegen müssen, doch wird er von vielen Autoren auch zum Einsatz bei Ordinaldaten vorgeschlagen, wenn gewisse Bedingungen erfüllt sind; es sollte $n > 20$ sein und die Anzahl r der Kategorien bei Ordinaldaten größer als 5 gewählt sein. Auch für metrische Daten gilt die Einschränkung, daß bei nur teilweise spezifizierter Verteilungshypothese unbedingt auf großen Stichprobenumfang und eine große Anzahl von Klassen geachtet werden muß. Falls explizite Parameter der Verteilung aus der Stichprobe geschätzt werden müssen, sollte bei kleinem Stichprobenumfang auch bei höherem Skalenniveau der χ^2 -Test angewandt werden (falls $n \cdot \pi_1 \geq 5$), dessen Prüfgröße dann eine Verteilung mit weniger Freiheitsgraden hat. Am besten wählt man den χ^2 -Test für alle Fälle mit nicht vollständig spezifizierter Verteilung, unabhängig von n .

Die Hypothese des Kolmogoroff-Smirnow-Tests lautet wie bei den beiden vorher genannten $H_0: F(x) = F_0(x)$. Ist diese zutreffend, so erwartet man für eine gegebene Stichprobengröße n eine Verteilungsfunktion $F_e(x)$ als Stichprobenrealisation, die tatsächlich beobachtete Verteilung

in der Stichprobe sei $F_b(x)$. Die verwendete Prüfgröße verwertet nun, anders als der χ^2 -Test auch die Anordnung der Werte, da sie auf den Abweichungen der beiden Verteilungsfunktionen F_b und F_e aufbaut, d. h. auf den kumulierten Häufigkeitsverteilungen. Als Maß für den Unterschied zwischen der erwarteten und der beobachteten Verteilung wird die betragsmäßig größte Differenz der entsprechenden kumulierten relativen Häufigkeiten gewählt, nämlich:

$$D = \max_{-\infty < x < \infty} |F_b(x) - F_e(x)|^{1)}.$$

Ist D klein, so

liegen die beiden Funktionen überall nah beieinander, ist D groß, so weichen sie zumindest an einer Stelle stark voneinander ab. Da immer $D \geq 0$ gilt, ist auch der Kolmogoroff-Smirnow-Test immer ein zweiseitiger Test mit einem Rückweisungspunkt.²⁾

Obwohl eine vollständig spezifizierte Verteilungsannahme für das Testverfahren nötig ist, zählt es zu den verteilungsfreien Verfahren, da die Verteilung der Prüfgröße D nicht von der in der Hypothese genannten Verteilungsfunktion F_0 bzw. von F_e abhängt, sondern nur vom Stichprobenumfang n. Jede der beiden Funktionen F_e und F_b kann nur Werte

-
- 1) Viele Autoren merken zu recht an, daß hier das Supremum statt des Maximums verwendet werden muß. Da jedoch in der Praxis die Spezifizierung " $-\infty < x < \infty$ " übertrieben ist, da es sich um das Maximum über alle tatsächlich in der Stichprobe aufgetretenen Werte handelt, wird hier auf das Supremum verzichtet.
 - 2) Für den einseitigen Test verwenden manche Autoren statt D die Prüfgrößen $D^+ = \max(F_e(x) - F_b(x))$ (für $H_a: F_b(x) < F_e(x)$) bzw. $D^- = \max(F_b(x) - F_e(x))$ (für $H_a: F_b(x) > F_e(x)$). Siehe z. B. BÜNING, H./TRENKLER, G.: Nichtparametrische Methoden, Berlin/New York, 1978

der Form $0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1$ annehmen, also auch der Betrag der Differenz; dessen Maximum ist ebenfalls einer der Werte a/n , $0 \leq a \leq n$ mit ganzzahligem a . Die Wahrscheinlichkeit, daß die Prüfgröße D einen bestimmten Wert $d = a/n$ erreicht oder überschreitet, bestimmt sich aus der Formel

$$P(D \geq d) = \frac{2 \cdot \left[\binom{2n}{n-a} - \binom{2n}{n-2a} + \binom{2n}{n-3a} - \dots \right]^{(1)}}{\binom{2n}{n}}$$

wobei die Reihe im Zähler fortgesetzt wird, solange $n-k \cdot a$ größer oder gleich n ist, d. h.

$$P(D \geq d) = \binom{2n}{n}^{-1} \cdot 2 \cdot \sum_{k=1}^q \binom{2n}{n-k \cdot a} \cdot (-1)^{k+1}$$

mit $q \cdot a \leq n$ und $(q+1) \cdot a > n$.

Für $n > 30$ kann die exakte Verteilung approximiert werden durch die Verteilung der Größe $Z = \sqrt{\frac{n}{2}} \cdot D$.²⁾

Zusammenfassung:

Anzahl der Stichproben	eine
Meßniveau	metrisch/ordinal (falls $n > 20$, $r \geq 5$)
Untersuchungsvariable	stetig
Hypothese	$F(x) = F_0(x)$
Alternative	zweiseitig
Prüfgröße	$D = \max_{-\infty < x < \infty} F_b(x) - F_e(x) $
Stichprobenverteilung	$P(D \geq d) = \binom{2n}{n}^{-1} \cdot 2 \cdot \sum_{k=1}^q \binom{2n}{n-k \cdot a} \cdot (-1)^{k+1}$ mit $q \cdot a \leq n$, $(q+1) \cdot a > n$

-
- 1) siehe dazu LEHMANN, E. L.: Nonparametrics. Statistical Methods Based on Ranks, San Francisco, 1975
 - 2) Vgl. SMIRNOW, N. Y.: Table for Estimating the Goodness of Fit of Empirical Distributions. In: Annals of Mathematical Statistics 19 (1948), S. 279-281

3.3 Prüfung auf Zufälligkeit-Iterationstest

Iterationstests haben die Aufgabe, die Zufälligkeit einer Abfolge von Zeichen zu überprüfen. Sie können somit benutzt werden, um zu klären, ob eine Stichprobe eine Zufallsstichprobe ist. Wenn in einer Reihe von n Werten nur 2 unterschiedliche auftreten können mit den Häufigkeiten n_1 und n_2 , so werden unter Umständen mehrere gleiche Werte aufeinander folgen. Eine solche Folge von gleichen Werten bezeichnet man als Run.¹⁾

Wenn die Abfolge der Werte zufällig zustande gekommen ist - dies ist die Hypothese eines Run-Tests -, so erwartet man eine "mittlere" Anzahl von Runs. Prüfgröße ist also die Variable X = Anzahl der Runs, deren Verteilung bei zutreffender Hypothese die folgende Wahrscheinlichkeitsfunktion hat:

$$P(X=x) = \frac{2}{\binom{n}{n_1}} \cdot \binom{n_1-1}{0,5x-1} \binom{n_2-1}{0,5x-1} \text{ für gerades } i,$$

$$P(X=x) = \frac{1}{\binom{n}{n_1}} \left[\binom{n_1-1}{x'-1} \binom{n_2-1}{x'-2} + \binom{n_1-1}{x'-2} \binom{n_2-1}{x'-1} \right] \text{ für ungerades } i,$$

wobei $x' = 0,5(x+1)$.

1) Deshalb werden solche Tests auch Run-Tests genannt.

Für $n_1, n_2 > 20$, also $n > 40$, läßt sich die Verteilung von x durch eine Normalverteilung approximieren mit dem Erwartungswert $\mu = \frac{2n_1n_2}{n}$ und der Varianz $\sigma^2 = \frac{2n_1n_2(2n_1n_2-n)}{n^2(n-1)}$.

Die exakte Verteilung von X wird für die Programme nicht benutzt; für $n < 50$ wird aber eine Stetigkeitskorrektur vorgenommen.

Die Alternative ist zweiseitig; sowohl "zu große" als auch "zu kleine" Werte von X führen zur Zurückweisung der Hypothese.

Aus der Konstruktion der Prüfgröße ist einsichtig, daß sich der Iterationstest auf beliebigem Skalenniveau einsetzen läßt, jedenfalls in der soeben beschriebenen elementaren Form.

Wird er hingegen zur Überprüfung der Zufälligkeit einer Stichprobenziehung benutzt, gilt dies nicht mehr. Die in der Stichprobe aufgetretenen Werte werden nämlich zunächst an ihrem Median dichotomisiert,¹⁾ damit nur die für den Run-Test erforderlichen zwei Werte vorliegen, bevor das Verfahren durchgeführt wird. Dazu ist also mindestens Ordinalskalenniveau nötig. Es sei denn, daß ohnehin ein dichotomes Merkmal vorliegt.

Wichtig für die Durchführung des Iterationstests ist die zeitliche Abfolge der positiven und negativen Abweichungen vom Median der Stichprobe; diese Abfolge muß bei der Stichprobenziehung festgehalten werden.

1) Bei metrischen Daten eventuell auch am arithmetischen Mittel.

Zusammenfassung:

Anzahl der Stichproben	eine
Meßniveau	nominal/ordinal (als Test für Stichproben)
Untersuchungsvariable	dichotom
Hypothese	Zufälligkeit der Abfolge
Alternative	zweiseitig
Prüfgröße	$X = \text{Anzahl der Runs}$
Stichprobenverteilung	$\sim N\left(1 + \frac{n_1 n_2}{n}; \frac{2n_1 n_2 (2n_1 n_2 - n)}{n^2 (n-1)}\right)$ (für $n_1, n_2 > 20$)

3.4 Prüfung von Unterschieden zwischen unabhängigen Stichproben

Die in diesem Abschnitt behandelten Tests werden eingesetzt zur Klärung der Frage, ob zwei oder mehr unabhängige Stichproben signifikante Unterschiede in der Verteilung aufweisen oder ob unterstellt werden kann, daß alle Stichproben aus derselben Population gezogen wurden. Auf diese Weise sollen Informationen gewonnen werden, z. B. über zeitliche Veränderungen einer Population, über geographische Unterschiede bei entsprechenden Grundgesamtheiten o. ä.. Dazu werden im wesentlichen zwei Wege beschriftet. Tests auf Mittelwertunterschiede prüfen die Abweichungen eines dem Skalenniveau entsprechenden Maßes der zentralen Tendenz in den Stichproben, d. h. sie klären die Frage, ob die Differenz der Stichprobenmittelwerte so groß ist, daß verschiedene Grundgesamtheiten unterstellt werden können, d. h. unterschiedliche Verteilungen der Grundgesamtheiten. Für manche dieser Verfahren ist es wichtig, gleiche Streuungen der Grundgesamtheitsverteilungen annehmen zu können, weshalb auch Verfahren zur Prüfung von Streuungsunterschieden aufgenommen wurden.

Anpassungstests¹⁾ prüfen in allgemeinerer Weise, ob sich die Verteilungen der Stichproben zugrundeliegenden Populationen unterscheiden, also nicht nur in Bezug auf den Mittelwert.

1) In diesem Kontext werden sie oft auch Homogenitätstests genannt.

3.4.1 Test auf Anpassung der Verteilungen

Die hier zu nennenden Tests prüfen allgemein die Gleichheit von Verteilungen in zwei oder mehr Stichproben. Das Prinzip des Anpassungstests ist schon aus Abschnitt 3.2 bekannt, in dem es um die Anpassung einer empirischen Verteilung an eine theoretische Verteilung ging. In diesem Abschnitt geht es nun um die Frage, ob mehrere unabhängige Stichproben aus Populationen mit gleicher Verteilung stammen, also möglicherweise aus derselben Population. Solche Tests gibt es für nominales und für ordinales Meßniveau. Spezielle Tests für metrisches Skalenniveau sind nicht vorgesehen.

Fisher-Test

Fisher's exakter Test ist bei beliebigem Skalenniveau einzusetzen, falls eine dichotome Untersuchungsvariable vorliegt, deren Verteilungsfunktion in zwei Populationen untersucht werden soll. Die Verteilungshypothese lautet $H_0: F_1(x) = F_2(x)$. Dies kann bei dichotomen Variablen umformuliert werden zu $H_0: \pi_1 = \pi_2 = \pi$, da die Gleichheit der Verteilungsfunktion auf diese Weise vollständig in H_0 enthalten ist. Die Alternative kann ein- oder zweiseitig sein.

Es wird allgemein vereinbart, daß als Prüfgröße die kleinste der möglichen vier Häufigkeiten der beiden Stichproben gewählt wird. Ist X_1 die Häufigkeit

der einen Ausprägung, deren Anteilswert in den Grundgesamtheiten π beträgt, in beiden Stichproben, so ist $X_1 + X_2$ bei zutreffender Hypothese bekannt, denn dann muß gelten $n \cdot \pi = X_1 + X_2$ ($n = n_1 + n_2$).

Die vier Häufigkeiten $X_1, X_2, n_1 - X_1, n_2 - X_2$ sind also alle bekannt, wenn eine von ihnen bekannt ist.¹⁾ Man wählt nach allgemeiner Konvention als Prüfgröße $X = \min(X_1, X_2, n_1 - X_1, n_2 - X_2)$. Die Verteilung von X ist eine hypergeometrische Verteilung mit den Parametern $N = n, M = n_1$ und $n \cdot \pi$ bei zutreffender Hypothese, d. h.

$$P(X) = \frac{\binom{n_1}{x} \binom{n-n_1}{n\pi-x}}{\binom{n}{n\pi}} = \frac{\binom{n_1}{x} \binom{n_2}{n_2-x}}{\binom{n}{n_1}}.$$

Die Berechnung dieser Wahrscheinlichkeiten ist recht langwierig, weshalb der Fisher-Test vornehmlich bei kleinen Stichproben eingesetzt wird, wenn der χ^2 -Test, der im Anschluß erläutert wird, noch nicht angewandt werden darf. Wie bei allen diskreten Verteilungen besteht die Möglichkeit, daß α nicht erreicht wird, d. h. daß die Wahrscheinlichkeit für eine Realisation aus dem Rückweisungsbereich unter dem Signifikanzniveau bleibt.

Zusammenfassung:

Anzahl der Stichproben	zwei unabhängige
Meßniveau	nominal
Untersuchungsvariable	diskret-dichotom
Hypothese	$\pi_1 = \pi_2 = \pi$
Alternative	ein- oder zweiseitig

1) Denn es gilt: $X_2 = \sum_{i=1}^2 X_i - X_1, n_2 - X_2 = n - n_1 - (X - X_1)$.

Ist also X_1 bekannt, so lassen sich $n_1 - X_1, X_2, n_2 - X_2$ herleiten, da $\sum_{i=1}^2 X_i, n_1, n_2$ bekannt sind.

Prüfgröße	$X = \min (X_1, X_2, n_1 - X_1, n_2 - X_2)$
Stichprobenverteilung	$H (n, n_1, n \cdot \pi)$

χ^2 -Test für zwei und mehr Stichproben

Die Voraussetzungen dieses Tests sind schon bekannt: es braucht nur nominales Skalenniveau vorzuliegen. Damit die Verteilung der Prüfgröße hinreichend genau angenähert werden kann, sollten die erwarteten Häufigkeiten nicht kleiner als fünf sein. Die Anzahl der Stichproben sei k , die Anzahl der Kategorien der Untersuchungsvariablen sei r . Die Verteilungshypothese lautet $H_0: F_1(x) = \dots = F_k(x)$; behauptet wird also, daß die k Stichproben aus einer Population oder zumindest daß sie aus Populationen mit gleicher Verteilungsfunktion der Untersuchungsvariablen stammen. Dem entspricht bei diskreten, polytomen Variablen die Behauptung, daß $\pi_{ij} = \pi_j$; $i=1, \dots, k$; $j=1, \dots, r$. Gelegentlich wird in der Literatur unterschieden zwischen den beiden Formulierungen der Hypothese; man spricht einmal vom Anpassungstest und zum anderen von Homogenitätstest. Dies ist nicht sinnvoll, weil erstens das formale Vorgehen völlig analog ist und weil zweitens die Anteilswerte π_{ij} die Verteilungsfunktionen vollständig spezifizieren, wenn es sich um diskrete, polytome Untersuchungsvariablen handelt.

Wie auch schon im Einstichprobenfall, als die Anpassung an eine theoretische Verteilung geprüft wurde, werden mit Hilfe der $F_i(x)$ bzw. der π_{ij} die sogenannten erwarteten Häufigkeiten bestimmt,

d. h. die Häufigkeiten e_{ij} , die bei zutreffender Hypothese in der Stichprobe i für die Kategorie j zu erwarten sind. Sie lassen sich bestimmen aus den Randhäufigkeiten $n_i = \sum_{j=1}^r b_{ij}$ (= Anzahl der Elemente der i -ten Stichprobe) und $m_j = \sum_{i=1}^k b_{ij}$ (= Anzahl der Elemente der j -ten Kategorie in allen Stichproben) zu $e_{ij} = \frac{n_i \cdot m_j}{n}$. (Dabei ist $n = \sum_i n_i = \sum_j m_j$ und b_{ij} die beobachtete Häufigkeit der Kategorie j in der i -ten Stichprobe). Anders ausgedrückt: für jede Kategorie und jede Stichprobe ist $e_{ij} = n_i \cdot \frac{m_j}{n} = n_i \cdot \pi_j$ bei zutreffender Hypothese. Die Prüfgröße, die Pearson'sche Stichprobenfunktion, ist

$$U = \sum_{i=1}^k \sum_{j=1}^r \frac{(b_{ij} - e_{ij})^2}{e_{ij}} = \sum_i \sum_j \frac{(b_{ij} - n_i \pi_j)^2}{n_i \pi_j}, \text{ d. h.}$$

wie im Einstichprobenfall werden die "relativen" Abweichungen vom erwarteten Wert aufsummiert. Da die Richtung der Abweichungen wiederum nicht berücksichtigt ist, kann die Alternative des Tests nur zweiseitig sein - auch im Fall $k = 2$ und $r = 2$, wenn die Hypothese $\pi_1 = \pi_2 = \pi$ lautet. Die Verteilung der Prüfgröße U folgt bei hinreichend großen Stichprobenumfängen (allgemein wird $e_{ij} \geq 5$ verlangt) einer χ^2 -Verteilung mit $(r-1)(k-1)$ Freiheitsgraden.

Falls Parameter der in der Hypothese genannten Verteilungsfunktion $F(x)$ aus der Stichprobe geschätzt werden müssen, reduziert sich die Anzahl der Freiheitsgrade entsprechend.

Schon vorher, bei der Behandlung des Fisher-Tests wurde darauf hingewiesen, daß der χ^2 -Test auch für den sogenannten Vierfelderfall eingesetzt werden kann

und auch eingesetzt werden sollte, sobald dies zulässig ist, da bei großen Stichprobenumfängen n_1, n_2 die hypergeometrische Verteilung umständlich zu berechnen ist. Für $k=2$ und $r=2$ reicht die Beschränkung $e_{ij} \geq 5$ nicht aus. Die Prüfgröße U hat in der Notation des Fisher-Tests folgende Gestalt:

$$U = \frac{n(X_1(n_2 - X_2) - X_2(n_1 - X_1))^2}{(\sum X_i)(n - \sum X_i) n_1 n_2}.$$

Die Verteilung von U folgt erst ab $n > 50$ hinreichend genau einer $\chi^2(1)$ -Verteilung;¹⁾ für $20 \leq n \leq 50$ verwendet man die sogenannte YATES'sche Kontinuitätsberichtigung, so daß statt U die Prüfgröße

$$U_{\text{kor}} = \frac{n(|X_1(n_2 - X_2) - X_2(n_1 - X_1)| - \frac{n}{2})^2}{(\sum X_i)(n - \sum X_i) n_1 n_2}$$

benutzt wird, die nun ausreichend exakt $\chi^2(1)$ -verteilt ist, falls $e_{ij} \geq 4$.

Für $n < 20$ sollte der Fisher-Test beim Zweistichproben-Fall und einem dichotomen Merkmal eingesetzt werden.

Der χ^2 -Test ist ein Test für beliebiges Meßniveau ohne Voraussetzungen an die Grundgesamtheitsverteilungen und ist deshalb vielseitig einsetzbar. Er verwendet aber nur die Informationen über Häufigkeiten, so daß bei höherem Skalenniveau Informationen über den Datenbestand nicht ausgenutzt werden für die Testentscheidung.

1) Wegen der geringen Anzahl der Freiheitsgrade, die im Abschnitt über den Fisher-Test bewiesen wurde, ist das minimale n hier so groß.

Zusammenfassung:

Anzahl der Stichproben	zwei oder mehr unabhängige
Meßniveau	nominal
Untersuchungsvariable	diskret-polytom
Hypothese	$F_1(x) = \dots = F_k(x) = F(x)$
Alternative	zweiseitig

$$\text{Prüfgröße} \quad U = \sum_{i=1}^k \sum_{j=1}^r \frac{(B_{ij} - e_{ij})^2}{e_{ij}}$$

$$(\text{im Vierfelderfall: } U = \frac{n(X_1(n_2 - X_2) - X_2(n_1 - X_1))^2}{(\sum X_{1j})(n - \sum X_{1j})n_1n_2})$$

$$\text{bzw. f. } 20 \leq n \leq 50: U_{\text{korr}} = \frac{n \left(\frac{1}{2} X_1(n_2 - X_2) - X_2(n_1 - X_1) \right)! \left(\frac{n}{2} \right)^2}{(\sum X_{1j})(n - \sum X_{1j})n_1n_2}$$

$$\text{Stichprobenverteilung} \quad \sim \chi^2((r-1)(k-1))$$

Wald/Wolfowitz-Runtest

Das nach Wald und Wolfowitz benannte Verfahren dient zur Prüfung der Hypothese, ob zwei unabhängige Stichproben aus derselben Population stammen. Strenggenommen wird vorausgesetzt, daß die Untersuchungsvariable stetig verteilt ist. Verstöße gegen diese Voraussetzungen haben aber keine gravierenden Auswirkungen. Der Test ist anwendbar auf Ordinalskalenniveau, da zur Prüfung nur die Rangzahlen der Stichprobenwerte herangezogen werden. Zunächst einmal werden die Werte der gepoolten Stichprobe der Größe nach geordnet. Das weitere Vorgehen basiert auf der Iterationstheorie, die schon beim Run-Test aus 3.3 angesprochen wurde. Allen Werten aus Stichprobe 1 wird nun ein bestimmtes Symbol ("+" oder 1) zugeordnet,

denen aus Stichprobe 2 ein zweites ("-" oder 2). Nun wird die Abfolge der Symbole in der der Größe nach geordneten Reihe der gepoolten Stichprobe untersucht. Stammen beide Stichproben aus derselben Grundgesamtheit, so ist eine große Anzahl X von Runs zu erwarten, weil die Symbole oft abwechseln werden.¹⁾ Eine kleine Anzahl von Runs kann bedeuten, daß die Grundgesamtheitsmediane sich unterscheiden oder daß - bei gleichen Medianen - die Streuung in den Populationen unterschiedlich ist. In jedem Fall ist bei einem genügend kleinen Wert von X auf zwei unterschiedliche Populationen zu schließen. Die Alternative lautet also stets auf Unterschiedlichkeit der Populationen. Schon in 3.3 wurde die Verteilung der Prüfgröße X genannt, deshalb soll hier darauf verzichtet werden, ebenso wie auf die Angabe der asymptotischen Verteilung für $n_1, n_2 > 20$. Verbundwerte für die Rangplatzdaten können bei exakt eingehaltener Stetigkeitsvoraussetzung nicht auftreten; da diese Bedingung aber in der Praxis oft nicht eingehalten ist, sind möglicherweise mehrere Werte für X möglich. Stammt nämlich ein Verbundwert aus beiden Stichproben, so ist die Abfolge der Symbole an dieser Stelle der Reihe nicht eindeutig. Je nach Wahl der Folge kann eine kleinere oder größere Zahl von X als Stichprobenrealisation auftreten. In Grenzfällen ist es durchaus möglich, daß dadurch die Testentscheidung von Ablehnung zu Annahme wechselt. Bei einer relativ großen Anzahl von Verbundwerten aus beiden Gruppen sollte der Wald/Wolfowitz-Runtest deshalb nicht

1) Man könnte sagen, die kleinen und großen Grundgesamtwerte sind gleichmäßig auf die beiden Stichproben verteilt.

verwendet werden.¹⁾ Er eignet sich ohnehin nur für Fälle, in denen jegliche Unterschiede zwischen den Grundgesamtheiten festgestellt werden sollten. Spezifische Unterschiede bezüglich Lokation oder Dispersion werden durch andere Verfahren besser geprüft. Der Test verwendet die Rangplatzinformationen, um die Abfolge der Symbole zu bestimmen. Seine Effizienz bei metrischen Daten gegenüber dem t-Test scheint gering zu sein,²⁾ was nicht verwundert, da der Wald/Wolfowitz-Test wenig Informationen aus den Daten verwendet und zudem nicht speziell auf Mittelwertunterschiede prüft, wie es der t-Test tut. Insgesamt muß das Verfahren als auch dem nachfolgend besprochenen Kolmogoroff/Smirnow-Test unterlegen angesehen werden. Bei kleinen Stichproben (etwa $n \leq 20$) ist seine Effizienz mit 75 % am höchsten, aber immer noch weitaus niedriger als die des Kolmogoroff/Smirnow-Tests (ca. 96 % bei kleinem n).

Zusammenfassung:

Anzahl der Stichproben	zwei unabhängige
Meßniveau	ordinal
Untersuchungsvariable	stetig
Hypothese	Die Populationen sind gleich
Alternative	zweiseitig
Prüfgröße	X = Anzahl der Runs

-
- 1) In solchen Fällen kann die minimale bzw. maximale Anzahl der Runs bestimmt werden. Der Test wird dann zweimal durchgeführt, um zu sehen, ob Unterschiede in der Testentscheidung auftreten.
 - 2) Genauere Informationen fehlen. Die Schätzungen reichen von 20 bis 75 %, je nach Stichprobengröße.

Stichprobenverteilung $\sim N\left(1 + \frac{n_1 n_2}{n}; \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n)}{n^2 (n-1)}}\right)$

(für $n_1, n_2 > 20$)

Kolmogoroff/Smirnow-Anpassungstest für zwei Stichproben

Wie auch der χ^2 -Anpassungstest ist der Kolmogoroff/Smirnow-Test nicht nur für die Prüfung der Anpassung einer empirischen an eine theoretische Verteilung anwendbar. Das Verfahren läßt sich erweitern auf die Prüfung der Frage, ob zwei unabhängige Stichproben aus Populationen mit gleichen Verteilungsfunktionen der Untersuchungsvariablen stammen. Es ist bekannt, daß der Test sich strenggenommen nur bei metrischem Meßniveau einsetzen läßt, doch ist es allgemein üblich, ihn auch für Ordinaldaten anzuwenden. Die Untersuchungsvariable wird als stetig vorausgesetzt. Die Hypothese lautet auf Gleichheit der beiden Verteilungsfunktionen, also

$$H_0: F_1(x) = F_2(x) = F(x).$$

Als Prüfgröße für den zweiseitigen Test dient die Variable D, die analog zum Einstichprobenfall wie folgt definiert ist:

$D = \max_{-\infty < x < \infty} |F_{1b}(x) - F_{2b}(x)|, ^1$ wobei F_{ib} die beiden beobachteten Verteilungsfunktionen sind. Die Verteilung der Größe D ist nicht von der in der Hypothese genannten Funktion F(x) abhängig; die

1) Für den einseitigen Test ist $D^+ = \max(F_{1b}(x) - F_{2b}(x))$ bzw. $D^- = \max(F_{2b}(x) - F_{1b}(x))$ zu wählen.

exakte Verteilung ist bereits im Abschnitt 3.2 hergeleitet worden, sie kann etwa ab $n > 30$ approximiert werden, wie SMIRNOW (1948) gezeigt hat. Über den Einsatz des Kolmogoroff/Smirnow-Tests ist ähnliches zu sagen wie beim Einstichprobenfall. Er kann angewandt werden statt des χ^2 -Tests bei Ordinaldaten, falls $n > 20$ und die Anzahl der Kategorien r größer als 5 ist. Bei kleinerem Stichprobenumfang können der Wald/Wolfowitz-Test oder der χ^2 -Test verwendet werden. Letzterer allerdings nur, wenn $r > 2$ erfüllt ist, sonst muß der Fisher-Test eingesetzt werden. Alle diese Alternativen nutzen weniger Informationen des Datenbestandes aus als der Kolmogoroff/Smirnow-Test.

Zusammenfassung:

Anzahl der Stichproben	zwei, unabhängig
Meßniveau	"ordinal"
Untersuchungsvariable	stetig
Hypothese	$F_1(x) = F_2(x) = F(x)$
Alternative	ein- oder zweiseitig
Prüfgröße	$D = \max_{-\infty < x < \infty} F_{1b}(x) - F_{2n}(x) $ (bzw. $D^+ = \max(F_{1b}(x) - F_{2n}(x))$ bzw. $D^- = \max(F_{2n}(x) - F_{1b}(x))$ für die einseitigen Tests)
Stichprobenverteilung	$P(D \geq d) = \binom{2n}{n}^{-1} \cdot 2 \cdot \sum_{k=1}^q \binom{2n}{n-k \cdot a} (-1)^{k+1}$ mit $q \cdot a \leq n$, $(q+1) \cdot a > n$

3.4.2 Mittelwertunterschiede

Die hier zu nennenden Testverfahren prüfen im Falle ordinalen Meßniveaus die Abweichung der Mediane, im Fall metrischen Meßniveaus die Abweichung der arithmetischen Mittel. Dabei handelt es sich um Tests für zwei Stichproben und um solche, die Unterschiede zwischen mehr als zwei Stichproben prüfen. Für nominales Meßniveau sind keine Verfahren zur Prüfung auf Mittelwertunterschiede, also Unterschiede des Modus, enthalten.

Median-Test für 2 und mehr Stichproben

Für Daten auf ordinalem Meßniveau soll die Frage geklärt werden, ob die Mediane mehrerer Stichproben sich nur zufällig unterscheiden oder ob ihre Differenzen so groß sind, daß der Schluß erlaubt ist, die Stichproben stammen aus unterschiedlichen Grundgesamtheiten. Die Vorüberlegungen sind für 2 oder mehr Stichproben gleich, die Testverfahren unterscheiden sich. Die Hypothese behauptet die Gleichheit aller Grundgesamtheitsmediane, die Alternative ist zweiseitig, bei nur zwei Stichproben kann auch gegen eine einseitige Alternative getestet werden.

Ist die Hypothese zutreffend, so wird der vermutete gemeinsame Grundgesamtheitsmedian aus einer "gepoolten" Stichprobe, nämlich aus der Zusammenlegung aller k Stichproben, geschätzt. Wenn sich die k Populationen hinsichtlich ihres Medians nicht unterscheiden, so muß in allen Stichproben der Anteil der Werte, die größer sind als der "gepoolte" Median nahe bei 0,5 liegen, er darf nur zufällig von 0,5

abweichen. Sei mit X_i , $i=1, \dots, k$, die Anzahl der Werte oberhalb des Median bezeichnet,¹⁾ so lautet also die Hypothese in formaler Schreibweise:

$$H_0: E\left(\frac{X_1}{n_1}\right) = \dots = E\left(\frac{X_k}{n_k}\right) = 0,5 \text{ oder - noch anders -}$$

$$H_0: \pi_1 = \dots = \pi_k = 0,5.$$

Diese Hypothese ist aus dem Abschnitt 3.4.1 bekannt, es ist die Hypothese des χ^2 -Homogenitätstests bzw. für $k=2$ die des Fisher-Tests.

- a) Liegen zwei unabhängige Stichproben vor, wird man "Fisher's Exact Probability Test" wählen, wenn kleine Stichprobenumfänge gegeben sind. Bei großen n_1 , n_2 ist dieser Test etwas umständlich, da die hypergeometrische Verteilung Stichprobenverteilung ist. Für die einseitige Alternative kann nur Fisher's Test verwendet werden. Das Vorgehen wurde im letzten Abschnitt ausführlich besprochen, deshalb kann hier darauf verzichtet werden.

Zusammenfassung:

Anzahl der Stichproben	zwei, unabhängige
Meßniveau	ordinal
Untersuchungsvariable(n)	dichotomisiert am Median
Hypothese	$E\left(\frac{X_1}{n_1}\right) = E\left(\frac{X_2}{n_2}\right) = 0,5$
Alternative	ein- oder zweiseitig
Prüfgröße	$X = \min(X_1, X_2, n_1 - X_1, n_2 - X_2)$

1) Diese Dichotomisierung am Median ist schon bekannt vom Iterationstest, wenn er zur Prüfung der Zufälligkeit einer Stichprobe eingesetzt wird.

Stichprobenverteilung $H(N, M, n) = H(n, n_1, X_1 + X_2)$

b) Bei mehr als zwei unabhängigen Stichproben muß immer der χ^2 -Homogenitätstest eingesetzt werden. In diesem Fall ist die Alternative zweiseitig. Dieser nur bei großen Stichproben verwendbare Test¹⁾ wird meist auch bei nur zwei Stichproben angewendet, wenn die Voraussetzungen über den Stichprobenumfang erfüllt sind, damit die langwierige Berechnung der hypergeometrischen Verteilung vermieden wird. In diesem Zusammenhang können weitergehende Ausführungen entfallen, da dieser Test schon eingehend behandelt worden ist.

Zusammenfassung:

Anzahl der Stichproben	mehr als zwei, unabhängige
Meßniveau	ordinal
Untersuchungsvariable(n)	dichotomisiert am Median
Hypothese	$E\left(\frac{X_1}{n_1}\right) = \dots = E\left(\frac{X_k}{n_k}\right)$
Alternative	zweiseitig
Prüfgröße	$U = \sum_{i=1}^k \sum_{j=1}^2 \frac{(X_{ij} - n_{i \cdot} \cdot \pi_j)^2}{n_{i \cdot} \cdot \pi_j}$
Stichprobenverteilung	$\sim \chi^2(k-1)$

1) Vgl. dazu die ausführlichere Darstellung in 3.4.1

Mann/Whitney-U-Test

Dieser Test wird für Ordinaldaten vorgeschlagen, obwohl eine Voraussetzung für seinen Einsatz die Stetigkeit der Verteilungsfunktion der Untersuchungsvariablen ist. Gegen diese Bedingung wird in der Praxis oft verstoßen; in dem Fall werden meist Verbundwerte (ties) auftreten, d. h. eine Rangzahl wird zu mehreren Stichprobenelementen gehören. Schon hier soll darauf hingewiesen werden, daß dann eine Korrekturgröße berücksichtigt werden muß.

Die Hypothese des Tests besagt, daß die Verteilungen der Untersuchungsvariablen gleich sind, bis auf einen Lageparameter, d. h. $F_1(x) = F_2(x + \lambda)$, wobei meist $\lambda = 0$ gegen die ein- oder zweiseitige Alternative $\lambda \neq 0$ bzw. $\lambda > 0$ oder $\lambda < 0$ getestet wird.

Die Frage, ob die Stichprobenrealisationen dieser Hypothese widersprechen, wird wiederum über eine gepoolte Stichprobe geklärt. Diese wird der Größe nach geordnet und jedem Element wird ein Rangplatz zugewiesen. Nach einer allgemein anerkannten Konvention werden die Stichproben so numeriert, daß $n_1 \leq n_2$ ist. Jedes Element der ersten Stichprobe wird nun mit allen Elementen der zweiten Stichprobe daraufhin verglichen, ob es einen höheren Rangplatz in der gepoolten Stichprobe hat, d. h. es werden $n_1 \cdot n_2$ Vergleiche durchgeführt. Die Prüfgröße U ¹⁾

1) Nicht zu verwechseln mit der Pearson'schen Stichprobenfunktion U aus den χ^2 -Tests.

zählt, wie oft ein x_1 -Wert größer ist als ein x_2 -Wert, d. h. es ist

$$U = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} A_{ij}(x_{1i}, x_{2j}) \text{ mit } a_{ij} = 1 \text{ für } x_{1i} > x_{2j} \\ \text{und } a_{ij} = 0 \text{ sonst.}$$

Eine andere Möglichkeit, U zu bestimmen, bietet die folgende Formel:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1,$$

wenn R_1 die Summe aller Ränge der ersten Stichprobe ist. Die Herleitung dieser Formel findet sich z. B. bei LEHMANN (1975), S. 12.

Falls die Hypothese $\lambda=0$ zutrifft, die beiden Verteilungen F_1 und F_2 somit übereinstimmen, ist $E(U) = 0,5 \cdot n_1 n_2$, da zu erwarten ist, daß genauso oft ein x_1 -Wert größer ist als ein x_2 -Wert wie ein x_1 -Wert kleiner ist als ein x_2 -Wert. (Gleichheit kann nicht auftreten, wenn ties ausgeschlossen sein sollen.)

Die exakte Verteilung von U läßt sich mit Hilfe der Kombinatorik bestimmen. Dies Verfahren ist umständlich, weshalb meist die Tatsache benutzt wird, daß U asymptotisch normalverteilt ist mit dem Mittelwert $\mu=0,5 \cdot n_1 n_2$ und der Varianz

$$\sigma^2 = \frac{n_1 n_2 (n+1)}{12}. \text{ Für } n_2 \geq 20 \text{ ist diese Näherung als hin-}$$

reichend gut zu betrachten. Für sehr kleine Stichproben bietet sich der Mediantest als Alternative an, dessen Effizienz für kleine Stichproben sehr hoch ist, gemessen am parametrischen t-Test.

Falls Verbundwerte aufgetreten sind, ist die Varianz der Prüfgröße zu korrigieren durch einen Faktor, der sich aus der Anzahl der ties ergibt:

$$\sum_{i=1}^m T_i = \sum_{i=1}^m (t_i^3 - t_i); \text{ dabei ist } t_i \text{ die Häufigkeit}$$

eines der m aufgetretenen Verbundwerte. Die korrigierte Varianz beträgt dann:

$$\sigma_{\text{korr}}^2 = \frac{n_1 n_2}{12} \left(n+1 - \frac{1}{n(n-1)} \sum_{i=1}^m T_i \right) \text{ oder - anders ge-}$$

schrieben -

$$\sigma_{\text{korr}}^2 = \frac{n_1 n_2}{n(n-1)} \left(\frac{n^3 - n}{12} - \frac{1}{12} \cdot \sum_{i=1}^m T_i \right).$$

Auf diese Weise kann die Verteilung von U zurückgeführt werden auf die Standardnormalverteilung, die leicht zu handhaben ist. Falls die Stichprobenrealisation von U größer ist als $0,5 \cdot n_1 n_2$, so geht man über zur komplementären Variablen $U' = n_1 n_2 - U$, die die gleiche Verteilung hat wie die ursprüngliche Variable U. Der Mann/Whitney-U-Test verwertet mehr Informationen als der vorher beschriebene Median-Test, da nicht nur dichotomisierte Werte betrachtet werden, sondern die Rangplatzinformation direkt in die Prüfgröße eingeht. Oft wird deshalb der Median-Test als Schnelltest bezeichnet, der dem Mann/Whitney-U-Test vorgeschaltet werden kann. Der Median-Test hat schwächere Voraussetzungen, da Stetigkeit der Untersuchungsvariablen nicht vorausgesetzt wird. Ebenso fehlt die Bedingung $F_1(x) = F_2(x+\lambda)$.

Die Güte des Mann/Whitney-Verfahrens ist hoch, im Vergleich zum t-Test beträgt die Effizienz 95 %, d. h. wenn die Verteilungsannahme des t-Tests nicht hinreichend gesichert erscheint, kann ohne weiteres auf den Mann/Whitney-Test zurückgegriffen werden.

Zusammenfassung:

Anzahl der Stichproben	zwei, unabhängige
Meßniveau	ordinal
Untersuchungsvariable	stetig, $F_1(x) = F_2(x + \lambda)$
Hypothese	$\lambda = 0$
Alternative	ein- oder zweiseitig

Prüfgröße
$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

(mit R_1 Rangsumme der ersten Stichprobe)

Stichprobenverteilung
$$\sim N(0, 5 \cdot n_1 n_2; \sqrt{\frac{n_1 n_2 (n+1)}{12}})$$

(bzw. $\sim N(0, 5 n_1 n_2;$

$$\sqrt{\frac{n_1 n_2}{n(n-1)}} \left(\frac{n^3 - n}{12} - \frac{1}{12} \sum_{i=1}^m T_i^2 \right)$$

wenn ties vorliegen. Dabei ist $T_i = t_i^3 - t_i$ mit: t_i Häufigkeit des i-ten Verbundwertes)

Kruskal/Wallis-H-Test

Der Kruskal/Wallis-H-Test¹⁾ kann als Verallgemeinerung des vorher besprochenen Mann/Whitney-U-Tests angesehen werden, dem ja nur 2 unabhängige Stichproben zugrunde lagen. Falls die Anzahl der Stichproben k größer ist als 2, ist dieses Verfahren zu erweitern. Die Hypothese lautet nun, daß die Mediane der k Grundgesamtheiten, denen die Stichproben entstammen, gleich sind. Vorausgesetzt wird eine stetige Untersuchungsvariable. Das Vorgehen beim Kruskal/Wallis-H-Test ist ähnlich wie beim U-Test von Mann/Whitney. Es wird eine gepoolte Stichprobe gebildet, deren Elementen Rangzahlen zugeordnet werden. Für jede der k Stichproben wird nun die Summe der Rangzahlen R_i , $i=1, \dots, k$, gebildet. Bei zutreffender Nullhypothese dürfen die "mittleren" Ränge der Stichproben $\bar{R}_i = \frac{R_i}{n_i}$ nur zufällig voneinander abweichen. Die Hypothese lautet also: $H_0: \bar{R}_1 = \dots = \bar{R}_k = \frac{n+1}{2}$, $i=1, \dots, k$. Als Prüfgröße für solche Abweichungen dient die Variable

$$H = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(n+1) .$$

Sie ist folgendermaßen abzuleiten: Für die Entscheidung ausschlaggebend sind die Abweichungen der $\bar{R}_i$ von $(n+1)/2$, dem mittleren Rang der gepoolten

1) KRUSKAL, W. H./WALLIS, W. A.: Use of ranks on one-criterion variance analysis. In: Journal of the American Statistical Association, 47 (1952), S. 583-621

Stichprobe und zwar ihre Größe, aber nicht ihre Richtung; also betrachten wir

$(\frac{R_i}{n_i} - \frac{n+1}{2})^2$ bzw. die Summe aller dieser Ab-

weichungen. Diese werden gewichtet, um zu verhindern, daß die Nullhypothese einfach deshalb angenommen wird, weil die n_i sehr groß sind. Es ergibt sich die Prüfgröße:

$$\begin{aligned}
 H &= \frac{12}{n(n+1)} \sum_{i=1}^k n_i \left(\frac{R_i}{n_i} - \frac{n+1}{2} \right)^2 \\
 &= \frac{12}{n(n+1)} \sum_{i=1}^k n_i \left(\frac{R_i^2}{n_i} - 2 \cdot \frac{n+1}{2} \cdot \frac{R_i}{n_i} + \frac{(n+1)^2}{4} \right) \\
 &= \frac{12}{n(n+1)} \cdot \sum_{i=1}^k \frac{R_i^2}{n_i} - \frac{12}{n(n+1)} \cdot 2 \cdot \frac{n+1}{2} \cdot \sum_{i=1}^k \frac{R_i}{n_i} \cdot n_i \\
 &\quad + \frac{12}{n(n+1)} \cdot \frac{(n+1)^2}{4} \sum_{i=1}^k n_i \\
 &\quad \quad \quad = n \\
 &= \frac{12}{n(n+1)} \cdot \sum_{i=1}^k \frac{R_i^2}{n_i} - \frac{12}{n} \frac{n(n+1)}{2} + \frac{12(n+1)n}{4n} \\
 &= \frac{12}{n(n+1)} \cdot \sum_{i=1}^k \frac{R_i^2}{n_i} - 6(n+1) + 3(n+1) \\
 &= \frac{12}{n(n+1)} \cdot \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(n+1)
 \end{aligned}$$

Dies ist die allgemein übliche Schreibweise, die oben bereits eingeführt wurde. Die Verteilung der Prüfgröße ist schon bei $n_i \geq 5$ und $k \geq 4$ ausreichend gut durch eine χ^2 -Verteilung mit $k-1$ Freiheitsgraden zu approximieren. Falls Verbundwerte auftreten, d. h. die Voraussetzung der Stetigkeit verletzt ist, wie es oft in der Praxis der Fall sein wird, muß die Größe H korrigiert werden mit Hilfe des schon bekannten Faktors $\sum_{i=1}^m T_i = \sum_{i=1}^m (t_i^3 - t_i)$. H wird dann ersetzt durch:

$$H_{\text{kor}} = H \cdot \left(1 - \frac{1}{n^3 - n} \sum_{i=1}^m T_i\right)^{-1}.$$

Das auch als Kruskal/Wallis-Einweg-Rangvarianzanalyse bekannte Verfahren ist von höherer Güte als der entsprechende Median-Test. Seine Voraussetzungen sind allerdings stärker, da Stetigkeit verlangt wird. Wegen der besseren Ausnutzung der Stichprobeninformationen wird der Kruskal/Wallis-Test als aussagekräftiger angesehen als der Median-Test.

Verglichen mit dem parametrischen t -Test beträgt die Effizienz 95 %, so daß bei unsicherer Verteilungsannahme der H -Test vorzuziehen ist.

Zusammenfassung:

Anzahl der Stichproben	mehr als zwei unabhängige
Meßniveau	ordinal
Untersuchungsvariable	stetig
Hypothese	$\frac{R_1}{n_1} = \dots = \frac{R_k}{n_k} = \frac{n+1}{2}$
Alternative	zweiseitig

Prüfgröße

$$H = \frac{12}{n(n+1)} \cdot \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(n+1)$$

$$(\text{bzw. } H_{\text{korrr}} = H(1 - \frac{1}{3} \frac{\sum_{i=1}^m T_i}{n-n})^{-1})$$

wenn ties vorliegen. Dabei ist $T_i = t_i^3 - t_i$, t_i -Häufigkeit des i -ten Verbundwertes).

Stichprobenverteilung

$$\sim \chi^2(k-1)$$

(für $n_i \geq 5$, $k \geq 4$)

t-Test für zwei unabhängige Stichproben

Dieses Verfahren ist nur einzusetzen bei metrischem Meßniveau; denn geprüft wird die Frage, ob Unterschiede zwischen den beiden arithmetischen Mitteln der Stichproben signifikant sind für unterschiedliche Mittelwerte in den zugrundeliegenden Populationen. Der t-Test gehört zu den parametrischen Testverfahren, da unterstellt wird, die Untersuchungsvariable sei normalverteilt mit den Mittelwerten μ_1 und den Varianzen σ_1^2 . Die Hypothese lautet $H_0: \mu_1 - \mu_2 = \delta_0$, wobei meist $\delta_0 = 0$ behauptet wird, d. h. die Stichproben stammen aus Grundgesamtheiten mit gleichem Mittelwert. Die Alternative kann ein- oder zweiseitig sein. Es sind bei diesem Testproblem drei Fälle zu unterscheiden:

a) σ_1^2 und σ_2^2 sind bekannt; b) σ_1^2 , σ_2^2 sind unbekannt, aber es kann $\sigma_1^2 = \sigma_2^2$ vorausgesetzt werden; c) σ_1^2 , σ_2^2 sind unbekannt und es ist $\sigma_1^2 \neq \sigma_2^2$. In den Fällen b) und c) wird der t-Test eingesetzt, im Fall a) ein einfacher Mittelwertdifferenztest.

- a) Sind σ_1^2 und σ_2^2 bekannt, also die Varianzen der Grundgesamtheiten, so dient als Prüfgröße die Differenz der Stichprobenmittelwerte $D_{\bar{x}} = \bar{X}_1 - \bar{X}_2$. Dieser Fall ist unproblematisch, da $D_{\bar{x}}$ normalverteilt ist bei normalverteilter Grundgesamtheit nach $N(\delta_0; \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$.

Somit läßt sich der Annahmebereich des Tests leicht über die Standardnormalverteilung bestimmen. Falls die Normalverteilungsannahme nicht begründet ist, kann bei großem Stichprobenumfang (ungefähr $n > 30$) zumindest eine Näherung an die Normalverteilung angenommen werden, und zwar mit den soeben genannten Parametern. Bei kleinem Stichprobenumfang muß auf ein verteilungsfreies Verfahren ausgewichen werden, z. B. den Mann/Whitney-U-Test oder einen Randomisierungstest.

Zusammenfassung:

Anzahl der Stichproben	zwei unabhängige
Meßniveau	metrisch
Untersuchungsvariable	stetig, normalverteilt mit σ_1 bekannt
Hypothese	$\mu_1 - \mu_2 = \delta_0$ (meist: $=0$)
Alternative	ein- oder zweiseitig
Prüfgröße	$D_{\bar{x}} = \bar{X}_1 - \bar{X}_2$
Stichprobenverteilung	$N(\delta_0; \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$ $(\sim N(\delta_0; \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}), \text{ falls}$ die n_i hinreichend groß und die Voraussetzung für die Untersuchungsvariable nicht erfüllt ist.)

- b) Sind σ_1^2 und σ_2^2 unbekannt, aber die Annahme $\sigma_1^2 = \sigma_2^2$ sinnvoll, so muß die Streuung aus den Stichproben geschätzt werden. Als Prüfgröße dient nun die Variable

$$t_1 = \frac{\bar{x} - \delta_0}{\sqrt{(n_1-1)S_1^2 + (n_2-1)S_2^2} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2} (n_1 + n_2 - 2)}}$$

Sie läßt sich herleiten als Quotient der beiden Variablen

$$K = \frac{\bar{x} - \delta_0}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

die standardnormalverteilt ist, und $\frac{Y}{n_1 + n_2 - 2}$ wobei

$$Y = \frac{1}{\sigma^2} ((n_1-1) \cdot S_1^2 + (n_2-1) S_2^2)$$

$\chi^2(n_1+n_2-2)$ - verteilt ist. Somit folgt sie einer $t(n_1+n_2-2)$ -Verteilung. Deshalb heißt dieses Verfahren t-Test. Für große Stichprobenumfänge läßt sich die Verteilung von t_1 durch die Standardnormalverteilung hinreichend genau annähern. Bei kleinen Stichprobenumfängen ist die exakte t-Verteilung zur Bestimmung des Annahmebereichs zu wählen.

Bei nicht erfüllter Normalverteilungsbedingung sollten verteilungsfreie Verfahren wie der Randomisierungstest eingesetzt werden.¹⁾ Sie müssen angewendet werden im Fall kleiner Stichproben.

1) Auch der Mann/Whitney-U-Test kann eingesetzt werden, er verwendet jedoch nur die Rangordnungsinformation, nicht die Information über die Größe der Werte wie der Randomisierungstest.

Für große Stichproben kann auf eine Annäherung an die Normalverteilung ausgewichen werden, und zwar mit den Parametern

$$\mu = \delta_0 \quad \text{und} \quad \sigma^2 = S^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

Dabei ist S^2 die "gepoolte" Varianz der Stichproben, d. h.

$$S^2 = \frac{(n_1-1) \cdot S_1^2 + (n_2-1) \cdot S_2^2}{n_1 + n_2 - 2}.$$

Zusammenfassung:

Anzahl der Stichproben	zwei unabhängige
Meßniveau	metrisch
Untersuchungsvariable	stetig, normalverteilt mit $\sigma_1 = \sigma_2 = \sigma$ unbekannt
Hypothese	$\mu_1 - \mu_2 = \delta_0$ (meist = 0)
Alternative	ein- oder zweiseitig
Prüfgröße	$t_1 = \frac{D_{\bar{x}} - \delta_0}{\sqrt{(n_1-1)S_1^2 + (n_2-1)S_2^2}} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2} \cdot (n_1 + n_2 - 2)}$ (bzw. $D_{\bar{x}} = \bar{X}_1 - \bar{X}_2$, falls n_i groß und die Normalverteilungsbedingung nicht erfüllt)
Stichprobenverteilung	$t(n_1 + n_2 - 2)$ (für t_1)

$$(\text{bzw. } \sim N(\delta_0; \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2} \cdot \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}})$$

für $D_{\bar{x}}$)

- c) Falls σ_1^2 und σ_2^2 nicht nur unbekannt, sondern auch noch verschieden sind, gibt es nur Näherungsverfahren. Für dieses sogenannte BEHRENS-FISHER-Problem liegt noch keine exakte Lösung vor. Es muß die Streuung beider Grundgesamtheiten getrennt geschätzt werden.

Als Prüfgröße dient nun die Variable

$$t_2 = \frac{\bar{D}_x - \delta_0}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$$

deren exakte Verteilung nicht angegeben werden kann.

Man geht davon aus, daß die Verteilung von t_2 ungefähr einer t-Verteilung folgt, deren Anzahl von Freiheitsgraden recht kompliziert zu berechnen ist, nämlich durch

$$v = \left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2 \cdot \left[\frac{\left(\frac{S_1^2}{n_1} \right)^2}{\frac{S_1^2}{n_1} + 1} + \frac{\left(\frac{S_2^2}{n_2} \right)^2}{\frac{S_2^2}{n_2} + 1} \right]^{-1} - 2$$

Wie schon unter b) kann auch in diesem Fall die Standardnormalverteilung bei großen Stichprobenumfängen herangezogen werden, damit die umständliche Berechnung von v vermieden wird. Bei kleinen Stichprobenumfängen sollte die $t(v)$ -Verteilung eingesetzt werden.

Falls die Normalverteilungsbedingung verletzt ist oder als unsicher erscheint, gilt ähnliches wie in b): Bei großem Stichprobenumfang kann als Stichprobenverteilung eine asymptotische Normalverteilung mit $\mu = \delta_0$ und

$$\sigma = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

eingesetzt werden, bei kleinem Umfang ist auf verteilungsfreie Verfahren auszuweichen.

Zusammenfassung:

Anzahl der Stichproben	zwei unabhängige
Meßniveau	metrisch
Untersuchungsvariable	stetig, normalverteilt mit $\sigma_1^2 = \sigma_2^2$, σ_1 unbekannt
Hypothese	$\mu_1 - \mu_2 = \delta_0$ (meist = 0)
Alternative	ein- oder zweiseitig

Prüfgröße

$$t_2 = \frac{\bar{D}_x - \delta_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

(bzw. $\bar{D}_x$ bei großem n_i
und ohne Normalverteilung)

Stichprobenverteilung $t(v)$ (für t_2) mit

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{\left(\frac{s_1^2}{n_1} \right)^2}{n_1 + 1} + \frac{\left(\frac{s_2^2}{n_2} \right)^2}{n_2 + 1}} - 2$$

(bzw. $\sim N(\delta_0; \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}})$ f. $\bar{D}_x$)

Die in den Abschnitten a)-c) erläuterten Verfahren prüfen alle die Mittelwertunterschiede zwischen zwei Grundgesamtheiten. Test a) stellt den einfachsten Fall dar, einen Mittelwertdifferenzen-Test, manchmal auch Gauß-Zweistichproben-Test genannt. Die beiden anderen sind zwei Versionen des t-Tests. Beiden gemeinsam ist das Vorgehen bei großen bzw. kleinen Stichprobenumfängen und im Fall der nicht gesicherten Nor-

malverteilungsannahme.

Bei kleinen Stichprobenumfängen muß unter der Normalverteilungsannahme die t-Verteilung benutzt werden; ist die Annahme verletzt oder in Frage gestellt, ist ein verteilungsfreies Verfahren zu wählen.

Bei großen Stichprobenumfängen kann unter Normalverteilungsannahme als Approximation der t-Verteilung die Standardnormalverteilung benutzt werden; ist die Annahme nicht gesichert, kann auf die Prüfgröße $D_{\bar{x}}$ ausgewichen werden, die einer

$N(\mu_1 - \mu_2; \sigma)$ -Verteilung folgt mit

$$\sigma = \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2} \cdot \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2}} \quad (\text{Fall b})$$

bzw. $\sigma = \sqrt{S_1^2/n_1 + S_2^2/n_2}$ (Fall c), d.h. es kann

auch die Standardnormalverteilung benutzt werden. Die Frage der Gleichheit von σ_1 und σ_2 kann mit dem später noch zu erwähnenden Varianzquotiententest geklärt werden, der dem t-Test vorgeschaltet werden sollte, damit die richtige Version ausgewählt werden kann.

Randomisierungstest

Der schon mehrfach erwähnte Test kann auf metrischem Meßniveau eingesetzt werden, ohne daß eine Normalverteilungsannahme erfüllt sein muß und ohne daß homogene Varianzen vorliegen müssen. Er kann sinnvoll eingesetzt werden, wenn Mittel-

wertunterschiede zwischen zwei unabhängigen Stichproben mit kleinen n_i geprüft werden sollen. Seine Effizienz ist höher als die des Mann/Whitney-U-Tests, da alle Stichprobeninformationen auch ausgenutzt werden. Für große Stichproben ist die exakte Stichprobenverteilung aufwendig zu berechnen, doch kann unter gewissen Bedingungen eine recht gute Näherung an die t-Verteilung angenommen werden, wenn man nicht die Näherung an die Normalverteilung benutzt, wie sie im vorigen Abschnitt beschrieben wurde. Die Hypothese lautet - wie vorher -

$H_0: \mu_1 - \mu_2 = 0$. Die Alternative kann ein- oder zweiseitig sein. Bei Gleichheit der Mittelwerte wird angenommen, daß beide Stichproben aus derselben Population stammen, in diesem Fall ist aber die Aufteilung der Werte auf die zwei Stichproben als zufällig anzusehen. Jede der

$$\binom{n_1+n_2}{n_1}$$

Möglichkeiten, die Gruppen einzuteilen, ist als gleichwahrscheinlich anzusehen, wenn H_0 zutrifft. Als Ablehnungsbereich sind nun die "extremen" Ereignisse festzulegen, d. h. die Gruppenaufteilungen, bei denen die Summen der beiden Datenreihen die größten Differenzen aufweisen. Die Größe dieses Bereichs, d. h. die Anzahl der extremen Ereignisse,

wird durch $\alpha \cdot \binom{n_1+n_2}{n_1}$ festgelegt, wenn α das Signifikanzniveau ist. Beim einseitigen Test, z. B.

$H_a: \mu_1 < \mu_2$, werden alle jene Gruppenaufteilungen dem Ablehnungsbereich zugeordnet, für die

$\sum_{i=1}^{n_1} X_{1i} - \sum_{i=1}^{n_2} X_{2i}$ besonders groß ist (bzw. bei $H_a: \mu_1 > \mu_2$

besonders klein), beim zweiseitigen Test sind sowohl

die positiven als auch die negativen Extrema in den Ablehnungsbereich einzuordnen.

Wenn die tatsächliche Aufteilung auf die Stichproben eines der

$$\alpha \cdot \binom{n_1+n_2}{n_1}$$

extremen Ereignisse ist, wird die Hypothese abgelehnt.

Man sieht, daß der Rechenaufwand für diesen Test recht groß ist. Es müssen für alle

$$\binom{n_1+n_2}{n_1}$$

Möglichkeiten der Aufteilung die Differenzen zwischen

den Summen $\sum_{i=1}^{n_1} X_{1i}$ und $\sum_{i=1}^{n_2} X_{2i}$ gebildet werden, aus

diesen die $\alpha \cdot \binom{n_1+n_2}{n_1}$ extremen ausgewählt werden. Dann muß geprüft werden, ob die Stichprobenrealisation in diesen Bereich fällt.

Aus diesem Grund wurde dieser Test nicht in das System REP-TEST mit aufgenommen, trotz seiner schwachen Voraussetzungen, durch die er sich für den Bereich der empirischen Sozialforschung gut eignen würde.

Einfaktorielle Varianzanalyse

Mit der Varianzanalyse steht ein Verfahren zur Verfügung, auch für mehr als zwei Stichproben Mittelwertunterschiede zu prüfen. Verglichen werden sollen die arithmetischen Mittel einer Untersuchungsvariablen in k Stichproben; dabei wird

metrisches Skalenniveau vorausgesetzt und die Untersuchungsvariable soll in allen k Grundgesamtheiten normalverteilt mit Mittelwert μ_i und unbekanntem, aber gleichem σ . Sind diese Voraussetzungen verletzt, muß auf den Kruskal/Wallis-H-Test ausgewichen werden. Die Hypothese dieses von R. A. Fisher¹⁾ in den zwanziger Jahren entwickelten Verfahrens lautet $H_0: \mu_1 = \dots = \mu_k = \mu$, die Alternative ist "zweiseitig", das heißt hier: nicht alle μ_i sind gleich. Die Prüfgröße beruht auf einer Zerlegung der Variation innerhalb einer "gepoolten" Stichproben, der Quadratsummenzerlegung der Gesamtvariation

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2 .$$

Abweichungen der Werte X_{ij} vom Gesamtmittelwert der gepoolten Stichprobe können durch Fehlereinflüsse zustande gekommen sein, d. h. es sind zufällige Abweichungen, oder durch einen systematischen Einfluß, der auf unterschiedliche Mittelwerte in den Grundgesamtheiten zurückzuführen ist. Man spricht von Variation innerhalb der Stichproben und Variation zwischen den Stichproben; es ist nämlich:

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 + \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2 .$$

Variation innerhalb der Stichp.

"zufällig"

Variation zwischen den Stich.

"systematisch"

1) FISHER, R. A.: Statistical methods for research workers, Edinburgh/London 1941

Ist der systematische Einfluß, also die Variation zwischen den Stichproben, groß im Vergleich zum zufälligen Einfluß, der Variation innerhalb der Stichproben, so wird man auf Unterschiede zwischen den Erwartungswerten μ_i schließen. Als Prüfgröße dient also der "Variationsquotient", nämlich:

$$F = \frac{\sum_{i=1}^k n_i (\bar{X}_{i.} - \bar{X})^2 / (k-1)}{\sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i.})^2 / (n-k)}$$

Bei zutreffender Nullhypothese ist die Zählervariable $\chi^2(k-1)$ -verteilt, die Nennervariable $\chi^2(n-k)$ -verteilt. Da beide Variablen stochastisch unabhängig sind, ist der Quotient aus den beiden Variationskomponenten $F(v_1, v_2)$ -verteilt mit $v_1 = k-1$ und $v_2 = n-k$ Freiheitsgraden. Die Varianzanalyse ist also eigentlich eine Variationsanalyse, die keine Aussagen über Varianzen macht, sondern über Mittelwertdifferenzen.

Eine wesentliche Voraussetzung für die Anwendung dieses Verfahrens war die Annahme gleicher Varianzen in den Grundgesamtheiten.

Ob diese Bedingung erfüllt ist, sollte zuvor mit einem Test geklärt werden. Dafür bietet sich bei normalverteilten Grundgesamtheiten der BARTLETT-Test an, der noch besprochen wird.

Die Varianzanalyse kann auch bei nicht normalverteilter Population eingesetzt werden, da sie relativ "robust" ist gegen die Verletzung dieser Voraussetzung. In dem Fall kann z. B. der χ^2 -Test zur Überprüfung der Gleichheit der Grundgesamtheitsverteilungen herangezogen werden, der sich aber nicht nur

auf die Prüfung der Streuungen beschränkt.

Zusammenfassung:

Anzahl der Stichproben	mehr als zwei unabhängige
Meßniveau	metrisch
Untersuchungsvariable	stetig, normalverteilt
	mit $\sigma_1 = \sigma_2 = \dots = \sigma_k = \sigma$
Hypothese	$\mu_1 = \mu_2 = \dots = \mu_k$
Alternative	zweiseitig
Prüfgröße	$F = \frac{\sum n_i (\bar{X}_{i.} - \bar{X})^2 / (k-1)}{\sum \sum (X_{ij} - \bar{X}_{i.})^2 / (n-k)}$
Stichprobenverteilung	$F(k-1, n-k)$

3.4.3 Streuungsunterschiede

Im vorigen Abschnitt wurden Tests auf Streuungsunterschiede schon mehrfach erwähnt, bei den Tests auf Mittelwertunterschiede, die gleiche Streuungen in den Grundgesamtheiten voraussetzen. In dem Fall sollten Tests auf Streuungsunterschiede den Mittelwerttests vorgeschaltet werden, um zu klären, ob die wesentliche Voraussetzung der Streuungsgleichheit erfüllt ist. Die bekannteren Verfahren setzen meist metrisches Skalenniveau voraus und prüfen die Gleichheit der Varianz. Auf Ordinalskalenniveau läßt sich der Test von Siegel/Tukey einsetzen, der auf Rangplätzen einer gepoolten Stichprobe basiert, allerdings nicht auf Rangplätzen im üblichen Sinn. Vielmehr wird dem kleinsten Wert der Platz 1, dem größten der Platz 2, dem zweitgrößten der Platz 3, dem zweitkleinsten der Platz 4 usw. zugewiesen. Dieser Test ist aber leider nicht symmetrisch, d. h. er kann zu unterschiedlichen Ergebnissen führen, je nachdem, ob man dem kleinsten

oder dem größten Element den Platz 1 zuweist. Aus diesem Grund und aus Platzgründen wurde er in das System REP-TEST nicht aufgenommen. Der Varianzquotiententest und der Bartlett-Box-Test müssen dem t-Test bzw. der Varianzanalyse vorgeschaltet werden. Sie werden hier erwähnt, da sie eine wesentliche Voraussetzung des t-Tests bzw. der Varianzanalyse klären, im System sind sie nicht enthalten aus den gleichen Gründen, die für den t-Test geltend gemacht wurden.

Varianzquotiententest

Voraussetzungen für dieses Verfahren sind metrisches Skalenniveau und zwei normalverteilte Grundgesamtheiten, über die Mittelwerte wird nichts vorausgesetzt. Die Abweichung der beiden Streuungen wird gemessen durch den Quotienten aus den Varianzen - daher der Name des Tests - die Hypothese lautet also $H_0: \sigma_1^2 / \sigma_2^2 = C_0$, wobei meist $C_0 = 1$, die Homogenitätshypothese, behauptet wird. Die Alternative kann lauten $\sigma_1^2 / \sigma_2^2 \neq 1$ bzw. $\sigma_1^2 / \sigma_2^2 < 1$ oder $\sigma_1^2 / \sigma_2^2 > 1$. Bezeichnen S_1^2 und S_2^2 die bekannten erwartungstreuen Schätzfunktionen für die Varianzen σ_1^2 und σ_2^2 , so sind die Variablen

$$(n_i - 1) / \sigma_i^2 \cdot S_i^2 \quad \chi^2\text{-verteilt}$$

mit $n_i - 1$ Freiheitsgraden, der Quotient

$$\frac{S_1^2 / \frac{1}{2}}{\sigma_1^2} / \frac{S_2^2 / \frac{1}{2}}{\sigma_2^2} \quad \text{also F-verteilt. Bei zutreffender Hypothese}$$

ist $\sigma_1^2 = \sigma_2^2$ 1), die Prüfgröße $F = S_1^2 / S_2^2$ ist somit $F(n_1 - 1, n_2 - 1)$ -verteilt.

-
- 1) Falls $C_0 \neq 1$ ist, ist die Prüfgröße $F = S_1^2 / (S_2^2 / C_0)$.
An der Verteilung ändert sich nichts.

Es ist Konvention, daß bei der Bildung von F stets die größere Varianz im Zähler steht, d.

h. allgemein gilt:

$$F = \max \left(\frac{s_1^2}{s_2^2}, \frac{s_2^2}{s_1^2} \right),$$

somit ist F immer größer oder gleich eins.

Oft wird dieses Verfahren auch einfach F-Test genannt, dies ist jedoch verwirrend, da manche Autoren auch den Test der einfaktoriellen Varianzanalyse, deren Prüfgröße ebenfalls F-verteilt ist, mit diesem Namen bezeichnen.¹⁾ Ist die Normalverteilungsannahme verletzt, kann bei sehr großen Stichprobenumfängen²⁾ auf einen einfachen Varianzdifferenzentest zurückgegriffen werden, dessen Hypothese $H_0: \sigma_1 - \sigma_2 = \delta_0$ (meist=0) über die Prüfgröße $D_2 = S_1 - S_2$ getestet wird, die asymptotisch normalverteilt ist nach

$$N \left(0; \sqrt{\frac{(n_1 + n_2) \cdot \sigma^2}{2 \cdot n_1 \cdot n_2}} \right),$$

falls die Hypothese zutrifft mit $\delta_0 = 0$, also

$$\sigma_1 = \sigma_2 = \sigma.$$

Zusammenfassung:

Anzahl der Stichproben	zwei unabhängige
Meßniveau	metrisch
Untersuchungsvariable	stetig, normalverteilt
Hypothese	$\sigma_1^2 / \sigma_2^2 = C_0$ (meist =1)
Alternative	ein- oder zweiseitig

1) Wie überhaupt die Namensgebung für die verschiedenen Testverfahren nicht immer einheitlich ist. So paßt z. B. der Name Mediantest auch auf den Mann/Whitney-U-Test und den Kruskal/Wallis-H-Test, bezeichnet aber hier doch ein spezielles Verfahren.

2) Etwa $n_1, n_2 \geq 100$

Prüfgröße	$F = s_1^2 / s_2^2$
Stichprobenverteilung	$F(n_1 - 1, n_2 - 1)$

Bartlett-Test

Bei der Varianzanalyse werden die Mittelwertunterschiede von k Stichproben auf ihre Signifikanz geprüft, dabei werden gleiche Varianzen in allen k Grundgesamtheiten vorausgesetzt. Da k auch größer als zwei sein kann, läßt sich der Varianzquotiententest nicht einsetzen, es sei denn, es werden mehrere Tests über wechselnde Quotienten durchgeführt. Sobald ein Test zur Ablehnung der Hypothese der Gleichheit der verwendeten zwei Varianzen führt, kann die Hypothese auf Gleichheit aller Varianzen zurückgewiesen werden. Bei diesem multiplen Vergleich muß das Signifikanzniveau des einzelnen Tests reduziert werden, damit das Gesamtsignifikanzniveau α nicht überschritten wird, wenn mehrere Tests nacheinander durchgeführt werden. Als globaler Test für die Hypothese $H_0: \sigma_1^2 = \dots = \sigma_k^2 = \sigma^2$ bietet sich der Bartlett-Test an, dessen Anwendung ebenfalls normalverteilte Grundgesamtheiten voraussetzt. Die Alternative lautet: Mindestens ein σ_i^2 ist nicht gleich σ^2 . Die Prüfvariable B des Bartlett-Tests ist etwas kompliziert, sie folgt näherungsweise einer $\chi^2(k-1)$ -Verteilung und hat die Gestalt

$$B = \frac{1}{c} \cdot \left[(n-k) \cdot \log \left(\frac{1}{n-k} \cdot \sum_{i=1}^k (n_i - 1) \cdot s_i^2 \right) - \sum_{i=1}^k (n_i - 1) \cdot \log s_i^2 \right]$$

mit

$$c = 1 + \frac{1}{3(k-1)} \cdot \left(\sum_{i=1}^k \frac{1}{n_i-1} - \frac{1}{n-k} \right) \quad 1) .$$

Der von SPSS verwendete Bartlett/Box-Test arbeitet mit einer noch komplizierter aufgebauten Prüfgröße, die auf B aufbaut, deren Verteilung aber eine F-Verteilung ist. Dieser aufwendige Weg soll in dieser Arbeit nicht besprochen werden. Es muß jedoch darauf hingewiesen werden, daß $S_i^2 \neq 0$ für alle i gelten muß, da sonst $\log S_i^2$ nicht berechnet werden kann.

Es wurde schon vorher darauf hingewiesen, daß bei nicht-normalverteilter Population auf ein verteilungsfreies Verfahren, wie z. B. den χ^2 -Test, ausgewichen werden muß.

Zusammenfassung:

Anzahl der Stichproben	mehr als zwei unabhängige
Meßniveau	metrisch
Untersuchungsvariable	stetig, normalverteilt
Hypothese	$\sigma_1^2 = \dots = \sigma_k^2 = \sigma^2$
Alternative	Es ex. i mit: $\sigma_i^2 \neq \sigma^2$
Prüfgröße	B (siehe oben)
Stichprobenverteilung	$\sim \chi^2(k-1)$

1) $\log$ bezeichnet hier den Logarithmus zur Basis e.
Es wurde die übliche Computerschreibweise benutzt.

3.5 Prüfung von Unterschieden zwischen verbundenen Stichproben

Das Problem der verbundenen Stichproben stellt sich in der Praxis meist in Form von Meßwertreihen derselben Gruppe von Untersuchungselementen, die zu verschiedenen Zeitpunkten (Paneluntersuchung z. B.) und/oder unter verschiedenen Bedingungen beobachtet oder befragt wurden. Ein klassisches Beispiel sind medizinische Untersuchungen derselben Personen mit und ohne Belastung. Die im folgenden genannten Tests lassen sich wiederum in zwei Gruppen einordnen; es sind solche, die Mittelwertunterschiede prüfen und solche, die - allgemeiner - Unterschiede in den Häufigkeitsverteilungen prüfen, hier in den Anteilswerten, da nur nominales Meßniveau Voraussetzung ist.

3.5.1 Anteilswertdifferenzen

Die typische Situation, in der die beiden hier zu nennenden Testverfahren eingesetzt werden, kann beschrieben werden durch die Einteilung "vorher/nachher" oder durch die Frage nach den "Wechslern": Bei wievielen Untersuchungselementen wurde bei der zweiten Messung ein anderes Ergebnis verzeichnet als bei der ersten? Die Frage, ob die Anteilswerte der verschiedenen Merkmalsausprägungen sich verändert haben von einer Messung zur anderen, wird also geklärt über Anzahl der Wechsler. Die interessierende Variable muß dichotom sein, das Skalenniveau kann beliebig sein.

Mc Nemar-Test

Dieses Verfahren setzt zwei verbundene Stichproben voraus, d. h. z. B. zwei Befragungen derselben Personen zu verschiedenen Zeitpunkten in diesem Fall ist $n_1=n_2=n$, die beiden zu untersuchenden Anteilswerte sind p_1 und p_2 . Bezeichnet X_{ij} ($i=1,2$; $j=1,2$) die Anzahl der Untersuchungselemente, die bei der ersten Messung die Ausprägung i und bei der zweiten die Ausprägung j hatten, so sind p_1 und p_2 wie folgt zu berechnen: $P_1=(X_{11}+X_{12})/n$ und $P_2=(X_{11}+X_{21})/n$.¹⁾

Man sieht, daß Unterschiede zwischen P_1 und P_2 ausschließlich auf die Größe von X_{12} und X_{21} zurückzuführen sind, da X_{11} in beiden Anteilswerten auftaucht. Diese beiden Variablen stellen aber gerade die Anzahl der Wechsler dar, derjenigen Untersuchungselemente also, bei denen die beiden Messungen zu unterschiedlichen Ergebnissen führen. Wenn die Hypothese $H_0: \pi_1=\pi_2$ zutrifft, so können die Realisationen X_{12} und X_{21} nur zufällig voneinander abweichen, d. h. der Anteil, den die Wechsler von Ausprägung 1 zu Ausprägung 2 an der Gesamtzahl der Wechsler haben, beträgt gerade 0,5. Zu prüfen ist also die Frage, ob $p=X_{12}/(X_{12}+X_{21})$ signifikant von 0,5 abweicht. Dies entspricht der Fragestellung des Binomialtests, dem eine $B(n,\pi)$ -Verteilung zugrundeliegt, wobei nun $n=X_{12}+X_{21}$ und $\pi=0,5$ zu setzen ist. Bei großem n kann - wie schon in Abschnitt 3.2 beschrieben - die Binomialverteilung durch die

1) Die beiden anderen ergeben sich zwangsläufig.

Normalverteilung approximiert werden. Statt dessen wählt nun McNemar bei großem Stichprobenumfang einen anderen Weg, er führt einen χ^2 -Test durch. Bei zutreffender Hypothese sind für X_{12} und X_{21} die Werte $(X_{12}+X_{21})/2$ zu erwarten, die Pearson'sche Stichprobenfunktion hat also die folgende Gestalt:

$$U = \frac{(X_{12} - \frac{X_{12}+X_{21}}{2})^2 + (X_{21} - \frac{X_{12}+X_{21}}{2})^2}{\frac{X_{12} + X_{21}}{2}} .$$

Nach einigen Umformungen erhält man die einfache Formel

$$U = \frac{(X_{12} - X_{21})^2}{X_{12} + X_{21}} .$$

Falls $n=X_{12}+X_{21}$ hinreichend groß ist, folgt U asymptotisch einer $\chi^2(1)$ -Verteilung. Falls $n < 30$ ist, sollte zur Verbesserung der Approximation eine Korrektur von U berücksichtigt werden, die Yates'sche Kontinuitätsberichtigung. Die Prüfgröße lautet dann wie folgt:

$$U_{\text{kor}} = \frac{(|X_{12} - X_{21}| - 1)^2}{X_{12} + X_{21}} .$$

Der McNemar-Test ist ein voraussetzungsarmes Verfahren, das eine in der Praxis oft auftauchende Frage beantworten kann, nämlich die, ob bestimmte Ereignisse signifikante Veränderungen hervorrufen, so daß sich die Anteilswerte eines dichotomen Merk-

males vorher und nachher wesentlich unterscheiden.

Zusammenfassung:

Anzahl der Stichproben	zwei verbundene
Meßniveau	nominal
Untersuchungsvariable	diskret-dichotom
Hypothese	$\pi_1 = \pi_2$
Alternative	zweiseitig ¹⁾
Prüfgröße	$U = \frac{(x_{12} - x_{21})^2}{x_{12} + x_{21}}$

$$\text{(bzw. } U = \frac{(|x_{12} - x_{21}| - 1)^2}{x_{12} + x_{21}} \text{)}$$

Stichprobenverteilung $\sim \chi^2(1)$ (für $n \geq 20$)

Cochran-Q-Test

Der McNemar ist nur für den Fall zweier verbundener Stichproben einzusetzen, eine Erweiterung auf k Stichproben leistet Cochran's-Q-Test. Die Untersuchungsvariable ist dichotom und es ist wiederum nominales Skalenniveau erlaubt. Die Hypothese lautet nun $H_0: \pi_1 = \dots = \pi_k = \pi$, wobei die π_i die Anteilswerte einer der beiden möglichen Ausprägungen sind zum Zeitpunkt i , bei Befragung i o. ä., allgemein: bei Messung i . Auch Cochran versucht die Frage der Gleichheit von Anteilswerten über die Zahl der Wechsler zu klären. Dies

1) Der exakte Binomialtest kann natürlich auch einseitig durchgeführt werden.

sind solche Untersuchungselemente, die nicht bei allen Messungen denselben Wert erreichen. Zu diesem Zweck verwendet er die etwas unübersichtliche Prüfgröße Q .¹⁾ Im Zähler befindet sich im wesentlichen die Streuung der Häufigkeiten einer der beiden Kategorien über die Stichproben hinweg; im Nenner hingegen steht eine Differenz von Summen über die Häufigkeiten der ausgewählten Kategorie bei jedem einzelnen Untersuchungselement.

Sei mit X_i die Häufigkeit der Kategorie 1 bei der i -ten Messung, mit L_j die Häufigkeit dieser Kategorie 1 beim j -ten Element bezeichnet, es sei n die Anzahl der Elemente, also der Stichprobenumfang, und k die Anzahl der Stichproben bzw. der Messungen. Die Prüfgröße Q des Cochran-Tests hat dann die folgende Gestalt:

$$Q = \frac{k(k-1) \cdot \sum_{i=1}^k (X_i - \bar{X})^2}{k \cdot \sum_{j=1}^n L_j - \sum_{j=1}^n L_j^2} \quad 1)$$

Falls $n \cdot k \geq 30$ gilt, folgt Q hinreichend genau einer χ^2 -Verteilung mit $k-1$ Freiheitsgraden.

Zusammenfassung:

Anzahl der Stichproben	k verbundene
Meßniveau	nominal

-
- 1) Die allgemeine Form dieser Prüfgröße wurde von FRIEDMAN entwickelt.
 - 2) Wichtig anzumerken ist die Tatsache, daß der Nenner von Q immer positiv ist bzw. nur dann gleich Null, wenn keine Wechsler auftreten, der Test also unsinnig ist. Über die Verteilung von Q und die allgemeine Form der Prüfgröße, vgl. LEHMANN (1975).

Untersuchungsvariable	diskret-dichotom
Hypothese	$\pi_1 = \dots = \pi_k = \pi$
Alternative	zweiseitig
Prüfgröße	Q (siehe oben)
Stichprobenverteilung	$\sim \chi^2(k-1)$ (für $n \cdot k \geq 30$)

3.5.2 Mittelwertdifferenzen

Wie schon im Fall unabhängiger Stichproben können auch für verbundene Stichproben nicht nur allgemein Unterschiede in den Verteilungen geprüft werden. Solche Testverfahren reagieren im allgemeinen auf jede Veränderung in den Verteilungen, woher sie auch kommen mag. Die im folgenden zu besprechenden Verfahren hingegen prüfen nur Unterschiede in den Mittelwerten, also in den Medianen bzw. in den arithmetischen Mitteln. Für nominales Meßniveau ist kein Test auf Unterschiede der Modi vorgesehen. Der t-Test ist nur für metrisches Skalenniveau vorgesehen, wobei zusätzlich wiederum die Normalverteilungsannahme erfüllt sein muß. Die anderen genannten Verfahren sind schon bei Ordinalniveau sinnvoll einzusetzen, der Friedmann-Test bei mehr als zwei Stichproben.

Dixon/Mood-Vorzeichen-Test

Dieses im Programmpaket SPSS kurz Sign-Test genannte Verfahren¹⁾ prüft den Unterschied zwischen

1) DIXON, W. J./Mood, A. M.: The statistical sign test.
In: JASA, Vol. 41 (1946), S. 557-566

den Medianen zweier verbundener Stichproben. Es wird die Unabhängigkeit der einzelnen Meßwertpaare vorausgesetzt und die Stetigkeit der betrachteten Untersuchungsvariablen gefordert. Auf das Problem der Stetigkeit von Verteilungen bei ordinalem Meßniveau ist schon mehrfach hingewiesen worden, es soll hier nur erwähnt werden. Die Hypothese H_0 lautet, daß die Mediane der beiden Meßwertreihen gleich sind. Wenn dies der Fall ist, so muß genau die Hälfte der Differenzen $X_{1i} - X_{2i}$ ¹⁾ zwischen den Meßwerten eines Paares positiv, die andere Hälfte negativ sein - von zufälligen Abweichungen abgesehen. Die Gleichheit der beiden Realisationen interessiert nicht, d. h. die Anzahl der zu untersuchenden Fälle reduziert sich um die Meßwertpaare, für die $X_{1i} - X_{2i} = 0$ ist. Prüfgröße ist das Minimum der Anzahl der positiven und der Anzahl der negativen Abweichungen, d. h. man setzt

$$V = \sum_{i=1}^n A_i \quad \text{mit} \quad \begin{aligned} A_i &= 1 && \text{für } d_i = x_{1i} - x_{2i} > 0 \quad \text{und} \\ A_i &= 0 && \text{für } d_i < 0 \quad , \end{aligned}$$

wobei n nicht die Gesamtzahl der Fälle ist, sondern die Anzahl der Fälle mit $d_i \neq 0$.

Weiter ist $V' = \sum_{i=1}^n (1 - A_i)$ und die Prüfgröße ist $\min(V, V')$.²⁾

Die Hypothese lautet nun formalisiert: $H_0: E(V) = 0,5 n$, die Alternative kann einseitig oder zweiseitig sein.

-
- 1) Natürlich ist bei ordinalem Skalenniveau die Bildung einer Differenz nicht erlaubt. Gemeint ist aber hier: Der Wert des ersten Elementes ist größer, kleiner oder gleich dem des zweiten. Da späterhin nur das Vorzeichen der Abweichung, nicht ihre Größe berücksichtigt wird, handelt es sich höchstens um einen Fehler in der Schreibweise.
 - 2) Man wählt das Minimum ausschließlich wegen der einfacheren Berechnung der Wahrscheinlichkeit, d. h. im Hinblick auf zu erstellende Programme soll die Berechnung so wenig aufwendig wie möglich gehalten werden, damit sich keine zu hohen Rechenzeiten ergeben. Im folgenden wird weiterhin mit V gearbeitet.

Bei zutreffender Hypothese ist die Variable V binomialverteilt nach $B(n; 0,5)$; diese Verteilung kann für $n > 25$ durch eine Normalverteilung angenähert werden, wie schon bei der Behandlung des Binomialtests ausgeführt wurde.

Es handelt sich bei diesem Verfahren um einen einfachen, schnell durchzuführenden Test mit schwachen Voraussetzungen - die Verletzung der Stetigkeitsbedingung, z. B. durch ties, wird über die Konstruktion der Prüfgröße teilweise abgefangen. Bei Verwendung der Normalverteilung wird natürlich eine Stetigkeitskorrektur berücksichtigt.

Der Dixon/Mood-Vorzeichentest kann ebenso wie der Mediantest im Fall unabhängiger Stichproben als Schnelltest angesehen werden, da er nur einen kleinen Teil der vorhandenen Informationen auswertet. Wenn nicht nur eine Rangordnung innerhalb der Paare möglich ist, sondern generell ordinales Skalenniveau vorliegt oder gar metrisches, so sollte auf den nachfolgend beschriebenen Wilcoxon-Vorzeichen-Rangtest zurückgegriffen werden.

Zusammenfassung:

Anzahl der Stichproben	zwei verbundene
Meßniveau	ordinal
Untersuchungsvariable	stetig
Hypothese	$E(V) = 0,5 n$
Alternative	ein- oder zweiseitig
Prüfgröße	$V = \text{Anzahl der positiven Abweichungen}$
Stichprobenverteilung	$B(n; 0,5)$ $(\sim N(0,5 n; 0,5 \sqrt{n}) \text{ für } n > 25)$

Wilcoxon-Vorzeichen-Rangtest

Dieses Verfahren prüft dieselbe Hypothese H_0 wie das soeben beschriebene, verwendet aber den absoluten Wert der Differenzen $D_i = X_{1i} - X_{2i}$. Diese $\{D_i\}$ werden mit Rangplätzen versehen und es werden die Summe W der Rangplätze der negativen Abweichungen und die Summe W' der Rangplätze der positiven Abweichungen bestimmt; d. h. es wird nicht nur vorausgesetzt, daß für jedes Paar eine Rangordnung festgelegt werden kann. Dies muß auch für die Abweichungen möglich sein.¹⁾ W und W' ergeben sich ähnlich wie V und V' aus

$$W = \sum_{i=1}^n A_i \operatorname{Rg}(|d_i|), \quad A_i = 1 \text{ f. } d_i > 0, \\ A_i = 0 \text{ f. } d_i < 0,$$

und $W' = \sum_{i=1}^n (1 - A_i) \operatorname{Rg}(|d_i|)$ (n ist wiederum die Anzahl der $d_i \neq 0$.) Die Hypothese lautet $H_0: E(W) = 0,25 \cdot n(n+1)$.

Wenn nämlich genau die Hälfte der Meßwertpaare eine positive Differenz hat, so ist bei gleichem Median zu erwarten, daß die positiven Differenzen gleichmäßig auf hohe und niedrige Rangplätze verteilt sind, d. h. die Hälfte der Gesamtsumme, dies ist $0,5 \cdot n \cdot (n+1)$, entfällt auf positive Differenzen, die andere auf negative, so daß ein Erwartungswert von $0,25 \cdot n(n+1)$ für W die Hypothese der Gleichheit der Mediane beschreibt. Wie schon bei der Prüfgröße V des Dixon/Mood-Vorzeichentests, wählt man die kleinere der Realisationen von W bzw. W' , um über die Rückweisung der Hypothese zu entscheiden. Diese Konvention hat

1) Man muß sich darüber im klaren sein, daß hier strenggenommen die Grenze des bei Ordinalniveau Erlaubten überschritten wird, denn die X_{ji} sind nicht notwendigerweise Rangplätze. Es werden nicht nur Differenzen, sondern auch noch Beträge gebildet, die dann als Grundlage für Rangplätze dienen.

sich allgemein durchgesetzt zur Verkürzung der Berechnung. Die Bestimmung der exakten Verteilung von W ist etwas aufwendig. Es gibt 2^n Möglichkeiten, die n Rangplätze an positive und negative Differenzen zu verteilen; die Wahrscheinlichkeit der Realisation w der Rangsumme der positiven Differenzen ergibt sich aus der Anzahl der geordneten n -Tupel von Plus- und Minuszeichen, bei denen die Addition der Positionsnummern (diese entsprechen den Rängen) der Pluszeichen gerade w ergibt. Ein Beispiel soll dies verdeutlichen: Für $n=3$ sind $2^3=8$ verschiedene 3er-Tupel möglich:

$(+,+,+)$, $(+,+,-)$, $(+,-,+)$, $(- ,+,+)$
 $w=1+2+3=6$, $w=1+2=3$, $w=1+3=4$, $w=2+3=5$
 $(+,-,-)$, $(-,+,-)$, $(-,-,+)$, $(-,-,-)$
 $w=1$, $w=2$, $w=3$, $w=0$

Somit ist $P(W=3) = \frac{2}{8} = 0,25$ und $P(W=w) = 0,125$ für $w=0,1,2,4,5,6$.

Für $n > 25$ kann diese langwierig zu bestimmende Verteilung durch die Normalverteilung approximiert werden mit dem Mittelwert $\mu = 0,25 \cdot n \cdot (n+1)$ und der Varianz $\sigma^2 = n \cdot (n+1) \cdot (2n+1) / 24$.¹⁾

Die Hypothese H_0 wird verworfen, wenn der mittlere Rang der positiven Differenzen zu weit von $0,25 \cdot n(n+1)$ abweicht.

Der Wilcoxon-Vorzeichen-Rangtest hat mit 95 %

1) Ein Beweis dieser Verteilungskonvergenz findet sich z. B. bei LEHMANN (1975), S. 351.

eine recht hohe Effizienz gegenüber dem parametrischen t-Test bei relativ schwachen Voraussetzungen. Die kritischen Anmerkungen zu seinem Ansatz bei Ordinaldaten sind damit jedoch nicht widerlegt, da beim Vergleich mit dem t-Test metrische Daten benutzt werden.

Zusammenfassung:

Anzahl der Stichproben	zwei verbundene
Meßniveau	ordinal
Untersuchungsvariable	stetig
Hypothese	$E(W) = 0,25 \cdot n(n+1)$
Alternative	zweiseitig
Prüfgröße	$W = \text{Rangsumme der positiven Differenzen}$
Stichprobenverteilung	$\sim N(0,25 \cdot n(n+1); \sqrt{n(n+1)(2n+1)/24})$

Friedman-Test

Da die beiden vorher genannten Verfahren die Differenzen d_j von Meßwertpaaren zur Konstruktion der Prüfgrößen V bzw. W benutzen, sind sie nur für zwei Stichproben einzusetzen. Falls mehr als zwei Messungen pro Fall durchgeführt wurden oder - allgemeiner gesagt - mehr als zwei homogenisierte Gruppen zu überprüfen sind, kann für den Test auf Abweichung der Mittelwerte das Friedman-Verfahren¹⁾ gewählt werden. Die Anzahl der Meßwertreihen werde allgemein mit k bezeichnet. Die "Friedman - Zweiweg-Rangvarianzanalyse" - so die ausführlichere Bezeichnung - weist Analogien auf zum Kruskal/Wallis-H-Test und zum Wilcoxon-Vor-

1) FRIEDMAN, M.: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. In: Journal of the American Statistical Association, Vol. 32 (1937), S. 675-701.

zeichen-Rangtest. Die Untersuchungsvariable wird als stetig angenommen und muß mindestens ordinalskaliert sein. Zu prüfen ist die Frage, ob die Mediane der k Meßwertreihen sich so sehr unterscheiden, daß ihre Abweichungen nicht mehr als zufällig angesehen werden können. Hypothese ist also die Gleichheit der Grundgesamtmediane, die Alternative ist zweiseitig. Für jede der n Versuchspersonen werden alle k Meßergebnisse¹⁾ in eine Rangordnung gebracht und mit Rangplätzen versehen. Sodann werden für jede der k Bedingungen die Ränge aller Personen addiert, so daß

$$R_i = \sum_{j=1}^n R_{ij}(X_{ij}), i=1, \dots, k, ^{2)}$$
, wenn X_{ij} die Realisation der Untersuchungsvariablen bei Person j unter Bedingung i darstellt. Falls nun die Hypothese zutrifft, daß die Mediane der k Meßwertreihen gleich sind, so ist zu erwarten, daß die Rangsummen R_i dieser k Reihen nur zufällig voneinander abweichen. Die Prüfung dieser Frage erfolgt nun sehr ähnlich wie beim Kruskal/Wallis-H-Test. Die Summe aller k Rangzahlen aller n Personen beträgt $n \cdot k(k+1)/2$. Wenn die Rangsummen der Meßwertreihen gleich sind, so ist $R_i = n(k+1)/2$ zu erwarten für alle i . Es wird untersucht, wie stark die R_i von diesem zu erwartenden Wert abweichen (die Richtung interessiert dabei nicht), d. h. man untersucht die Größen: $(R_i - n(k+1)/2)^2$ bzw. die Summe dieser Werte, so ergibt sich eine

-
- 1) Der Einfachheit halber soll angenommen werden, daß n Versuchspersonen unter k verschiedenen Bedingungen untersucht werden.
 - 2) Man beachte den Unterschied zum Kruskal/Wallis-Test, bei dem R_i die Summe der Rangzahlen in der gepoolten Stichprobe für eine Gruppe i war. Hier hingegen sind die Ränge bezogen auf die einzelne Person.

ähnliche Prüfgröße wie H, nämlich

$$FR = \frac{12}{n \cdot k(k+1)} \cdot \sum_{i=1}^k \left(R_i - \frac{n(k+1)}{2} \right)^2 \text{ ergibt.}$$

Analog zum Vorgehen bei der Einweg-Rangvarianzanalyse kann man sie umformen zu der üblichen Schreibweise:

$$FR = \frac{12}{n \cdot k(k+1)} \cdot \sum_{i=1}^k R_i^2 - 3 \cdot n(k+1) .$$

Diese Prüfgröße ist schon bei relativ kleinen Stichprobenumfängen (für $k=3$ ab $n=10$, für $k=4$ ab $n=5$) recht gut durch eine χ^2 -Verteilung mit $k-1$ Freiheitsgraden zu approximieren.

Die Voraussetzungen des Friedman-Tests sind relativ schwach, es wird lediglich gefordert, daß sich die Meßergebnisse einer Versuchsperson in eine geordnete Reihe bringen lassen. Die Stetigkeit der Untersuchungsvariablen kann - wie so oft - in vielen Anwendungsfällen nicht angenommen werden. Treten in den Rängen einer Person ties auf, so verschieben diese das Testergebnis möglicherweise, es ist aber nicht zu klären, in welche Richtung. Genauere Angaben über die Effizienz des Friedman-Tests liegen nicht vor, doch ist zu vermuten, daß ein Vergleich mit der metrischen Zweiweg-Varianzanalyse, die hier nicht besprochen wird, sehr gut ausfallen wird, daß möglicherweise sogar nicht entschieden werden kann, welches der beiden Verfahren der stärkere Test ist.

Zusammenfassung:

Anzahl der Stichproben	k verbundene
Meßniveau	ordinal
Untersuchungsvariable	stetig
Hypothese	Gleichheit aller k Mediane
Alternative	zweiseitig
Prüfgröße	$FR = \frac{12}{nk(k+1)} \cdot \sum_{i=1}^k R_i^2 - 3n(k+1) ,$ wobei R_i die Summe der Ränge aller Personen bei Messung i ist.
Stichprobenverteilung	$\sim \chi^2(k-1)$

t-Test für zwei verbundene Stichproben

Dieses Verfahren soll kurz erwähnt werden, da es zur Bestimmung der Effizienz des Dixon/Mood- und des Wilcoxon-Tests herangezogen wird. Das Vorgehen ist ähnlich wie beim t-Test für unabhängige Stichproben und soll deshalb nicht ausführlich erläutert werden, zumal es nicht im System REP-TEST enthalten ist. Man betrachtet die Variable $D=X_1-X_2$ bzw. ihr arithmetisches Mittel $\bar{D}=\bar{X}_1-\bar{X}_2$.

Bei großen Stichprobenumfängen ist $\bar{D}$ approximativ normalverteilt. Falls die Hypothese $E(\bar{D})=\delta_0$ zutrifft (meist ist dabei $\delta_0=0$), ist $\mu=\delta_0$ und

$$\sigma^2 = \frac{1}{n(n-1)} \cdot \sum_{i=1}^n (D_i - \bar{D})^2 .$$

Bei kleinen Stichprobenumfängen muß die Voraussetzung der Normalverteilung von D erfüllt sein.

Bei unbekannter Varianz ist die Prüfgröße

$$\frac{\bar{D} - \delta_0}{\sqrt{\frac{1}{n(n-1)} \cdot \sum (D_i - \bar{D})^2}} \quad t(n-1)\text{-verteilt.}$$

Ist die Normalverteilungsannahme nicht berechtigt, so wird ein verteilungsfreies Verfahren, z. B. der Wilcoxon-Test, eingesetzt.

Zusammensetzung:

Anzahl der Stichproben zwei verbundene

Meßniveau metrisch

Untersuchungsvariable stetig

Hypothese $\mu_1 - \mu_2 = \delta_0$ (meist = 0)

Alternative ein- oder zweiseitig

Prüfgröße $\bar{D}$ (für große n)

bzw. $(\bar{D} - \delta_0) / \sqrt{\frac{1}{n(n-1)} \cdot \sum (D_i - \bar{D})^2}$

für kleine n, wenn $\bar{D}$ normalverteilt ist.

Stichprobenverteilung $N(\delta_0; \sqrt{\frac{1}{n(n-1)} \cdot \sum (D_i - \bar{D})^2})$

bzw. $t(n-1)$

3.6 Prüfung des Zusammenhangs zwischen Merkmalen

Ein letzter Problemtyp, der in der Praxis häufig auftaucht, soll noch besprochen werden. Die Frage, ob zwischen zwei Merkmalen ein Zusammenhang besteht oder nicht, ob die Größe der Werte einer Variablen die einer zweiten beeinflusst, stellt sich oft auch schon im Bereich der deskriptiven Statistik. Es werden dann auf den unterschiedlichen Skalenniveaus Zusammenhangsmaße bestimmt, auch Assoziationsmaße oder Korrelationsmaße genannt. Für metrische Datenqualität ist der Korrelationskoeffizient von Bravais/Pearson der bekannteste. Für ordinales Meßniveau sind die Rangkorrelationskoeffizienten nach Spearman und Kendall am gebräuchlichsten. Für REP-TEST wurde der Spearman'sche Koeffizient gewählt. Eine Fülle von Assoziationsmaßen sind für Nominaldaten bekannt; eine Gruppe ist die der χ^2 -basierten Maße, zu der λ , ϕ und C gehören. Ihre Überprüfung erfolgt mit dem χ^2 -Unabhängigkeitstest. Die Hypothese lautet im allgemeinen auf Unabhängigkeit der Merkmale, d. h. es wird behauptet, das entsprechende Zusammenhangsmaß sei gleich Null.

3.6.1 χ^2 -Unabhängigkeitstest

Das Vorgehen beim Unabhängigkeitstest ähnelt sehr dem des χ^2 -Homogenitätstests, der in 3.4.1 beschrieben wurde; auch die Prüfgröße hat eine ähnliche Gestalt, wenn auch die Fragestellung eine andere ist. Betrachtet werden zwei Variablen beliebigen Meßniveaus, die diskret sind. Ob dichotom oder polytom ist zunächst nicht von Bedeutung. Es sei B_{ij} die Häufigkeit, mit der die Kombination der Kategorie i bei der einen Variablen (X) und der Kategorie j bei der anderen (Y) aufgetreten ist. Die Randhäufigkeiten n_i bzw. m_j für die beiden Variablen bestimmen zusammen mit dem Stichproben-

umfang n die für die Prüfgröße benötigten erwarteten Häufigkeiten. Falls nämlich die zu untersuchenden Merkmale unabhängig sind, wie es die Hypothese behauptet, so gilt für die Häufigkeit einer Kombination (X_i, Y_j) : Die n_i Fälle, die die Ausprägung X_i haben, verteilen sich auf alle Ausprägungen von Y entsprechend den Gewichten, die die Ausprägungen Y_j in der Stichprobe haben; dies ist gerade m_j/n , wenn n der Stichprobenumfang ist. Also ist bei Unabhängigkeit eine Häufigkeit $e_{ij} = n_i \cdot m_j / n$ zu erwarten. Als Prüfgröße dient wiederum die Pearson'sche Stichprobenfunktion U , die Summe der "prozentualen Abweichungen" der beobachteten von den erwarteten Häufigkeiten, nämlich

$$U = \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} \frac{(B_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} \frac{(B_{ij} - n_i m_j / n)^2}{n_i m_j / n},$$

wobei die k_i die Anzahlen der Ausprägungen beider Merkmale bezeichnen.

Die formale Analogie zum Homogenitätstest ist offensichtlich. Der Unterschied liegt in der Anzahl der Stichproben und der Untersuchungsvariablen: Der Unabhängigkeitstest prüft zwei Merkmale in einer Stichprobe, der Homogenitätstest dagegen ein Merkmal in zwei unabhängigen Stichproben.

Wegen der großen Ähnlichkeit kann jedoch auf weitere Ausführungen verzichtet werden. U folgt bekanntermaßen einer $\chi^2((k_1-1) \cdot (k_2-1))$ -Verteilung. Bei zwei dichotomen Merkmalen sollte die Yates'sche Korrektur berücksichtigt werden für $20 \leq n \leq 50$. Falls $n < 20$ gilt, ist der Fisher-Test, übertragen auf den

Fall zweier Merkmale, das geeignete Instrument.

Zusammenfassung:

Anzahl der Stichproben	eine
Meßniveau	nominal
Untersuchungsvariablen	zwei diskrete
Hypothese	$E(U)=0$
Alternative	zweiseitig
Prüfgröße	$U = \sum_i \sum_j \frac{(B_{ij} - n_{i.}m_{.j}/n)^2}{n_{i.}m_{.j}/n}$ $(\text{ bzw. } U = \frac{n (b_{11}b_{22} - b_{12}b_{21} - n/2)^2}{n_1 n_2 m_1 m_2}$ für 2 dichotome Merkmale und $20 \leq n \leq 50$)

Stichprobenverteilung $\sim \chi^2((k_1-1)(k_2-2))$

Rangkorrelationskoeffizient nach Spearman

Dieses Zusammenhangsmaß basiert auf den Differenzen von Rangplätzen eines Falles in beiden Merkmalen, $D_i = Rg(X_i) - Rg(Y_i)$, und entspricht dem Korrelationskoeffizienten von Bravais/Pearson, wenn dieser für die Rangplätze von metrischen Daten ohne ties berechnet wird.

Die Berechnungsformel lautet:

$$r_{SP} = 1 - \frac{6 \cdot \sum_{i=1}^n d_i^2}{n^3 - n}$$

Wenn kein Zusammenhang zwischen den Variablen X und Y besteht, ist $r_{SP}=0$. Dies ist die Hypothese des entsprechenden Unabhängigkeitstests -

$H_0: r_{SP}=0$ oder exakter $H_0: E(R_{SP})=0$.

Die exakte Verteilung von R_{SP} ¹⁾ ist etwas umständlich zu bestimmen: Zu jeder denkbaren Rangreihe der x_i gibt es $n!$ Möglichkeiten, die Ränge der y_i zu vergeben. Bei zutreffender Hypothese sind alle diese Möglichkeiten gleichwahrscheinlich, d. h. ihre Wahrscheinlichkeit beträgt $1/n!$. Alle Permutationen der Ränge der y_i , die zu demselben Wert für $\sum d_i^2$ führen, werden als günstige Fälle betrachtet; so wird Schritt für Schritt durch Auszählen die Wahrscheinlichkeit einer bestimmten Höhe von $\sum d_i^2$, also von r_{SP} , ausgerechnet. Ab $n > 10$ ist aber die Verteilung der Prüfgröße

$$R_{SP} \sqrt{\frac{n-2}{1-R_{SP}^2}}$$

hinreichend gut durch eine $t(n-2)$ -Verteilung anzunähern, was uns die langwierige Berechnung für die meisten Anwendungsfälle erspart. Für noch größere n , etwa $n > 20$ kann auch die noch einfacher zu berechnende Standardnormalverteilung benutzt werden, der die Größe $R_{SP} \sqrt{n-1}$ dann hinreichend genau folgt, da die t -Verteilung konvergent ist gegen die $N(0,1)$ -Verteilung.

Wichtig ist noch die Behandlung von ties. In diesem Fall sollte der Rangkorrelationskoeffizient korrigiert werden. Dazu benutzt man die schon früher eingeführte Größe $T = \sum_{i=1}^m (t_i^3 - t_i)$, wobei m die Anzahl der mehrfach aufgetretenen Werte ist

1) Vgl. HOTELLING, H./PABST, M. R.: Rank correlation and tests of significance involving assumption of normality. In: Annals of Mathematical Statistics, Vol. 7(1936), S. 29-43

und t_i die Häufigkeit des i -ten Verbundwertes. Hier müssen Korrekturen für ties in X und in Y berücksichtigt werden, dadurch verändert sich R_{SP} zu:

$$R_{SP} = \frac{T_X' + T_Y' - \sum_{i=1}^n d_i^2}{2 \cdot \sqrt{T_X' \cdot T_Y'}} \quad , \text{ dabei steht } T_X' \text{ für}$$

$$T_X' = \frac{N^3 - N \cdot T_X}{12} \quad , T_Y' \text{ wird analog}$$

bestimmt. T_X und T_Y sind die oben genannten Werte für die beiden Variablen. Falls keine ties auftreten, so sind T_X und T_Y gleich Null und R_{SP} hat die bekannte einfachere Form. Falls alle Werte mindestens einer Variablen gleich sind, ist die entsprechende Größe T gerade gleich $N^3 - N$, d. h. T' ist Null. In dem Fall kann R_{SP} nicht bestimmt werden; es wäre auch unsinnig, da ein Zusammenhang beider Merkmale nicht feststellbar ist, wenn eines nicht variiert.

Für den Fall eines vollkommenen Zusammenhangs der beiden Merkmale, wenn also R_{SP} die Werte -1 oder $+1$ annimmt, kann die Prüfgröße $t = R_{SP} \sqrt{\frac{n-2}{1-R_{SP}^2}}$

nicht berechnet werden. Ein solcher Wert wird aber sicherlich nicht mehr als zufällige Abweichung von Null angesehen werden.

Die Alternative wird meist zweiseitig formuliert, nämlich $r_{SP} \neq 0$, doch sind auch einseitige Alternativen möglich.

Die Anwendung des Tests für den Rangkorrelationskoeffizienten von Spearman erfordert nur Ordinalskalenniveau, weitere Voraussetzungen müssen nicht erfüllt sein, falls man nicht $|R_{SP}| < 1$ und $T_X, T_Y \neq N^3 - N$

als solche ansieht. Auch bei metrischen Daten kann er benutzt werden, wenn der Korrelationskoeffizient von Bravais/Pearson strenggenommen nicht eingesetzt werden darf, weil seine Berechnung auf der Annahme beruht, daß die Verteilung der Zufallsvariablen (X,Y) eine zweidimensionale Normalverteilung ist.

Zusammenfassung:

Anzahl der Stichproben	eine
Meßniveau	ordinal
Untersuchungsvariablen	zwei, diskret oder stetig
Hypothese	$E(R_{SP})=0$
Alternative	ein- oder zweiseitig
	$t = R_{SP} \sqrt{\frac{n-2}{1-R_{SP}^2}}$ mit
	$R_{SP} = \frac{T_X' + T_Y' - \sum d_i^2}{\sqrt{2 T_X' T_Y'}}$, wobei
	$T' = (N^3 - N - T)/12$ und $T = \sum (t_i^3 - t_i)$
Stichprobenverteilung	$\sim t(n-2)$ (für $n > 10$)

Korrelationskoeffizient von Bravais/Pearson

Dieses wohl bekannteste Zusammenhangsmaß, das deshalb auch oft nur als Korrelationskoeffizient bezeichnet wird, beruht auf den Varianzen der beiden betrachteten Untersuchungsvariablen und ihrer Kovarianz; die Formel lautet bekanntermaßen:

$$r = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} .$$

Somit ist klar, daß r nur für metrische Daten berechnet werden darf, also auch der entsprechende Test nur in diesem Fall angewandt werden darf. Schon bei der Behandlung des Spearman-Koeffizienten wurde darauf hingewiesen, daß die exakte Verteilung der Variablen R nur dann zu bestimmen ist, wenn (X,Y) einer zweidimensionalen Normalverteilung folgt. Falls nämlich ρ der Korrelationskoeffizient der Grundgesamtheit ist, so folgt die Stichprobenvariable $R^{1)}$ einer Verteilung mit der folgenden Dichtefunktion f :

$$f(r) = \frac{n-2}{\pi} \cdot (1-\rho^2)^{\frac{1}{2}(n-1)} \cdot (1-r^2)^{\frac{1}{2}(n-4)} \int_0^1 \frac{\xi^{n-2}}{(1-\rho r \xi)^{n-1}} \cdot \frac{1}{\sqrt{1-\xi^2}} d\xi.$$

Für gegebenen Stichprobenumfang n hängt $f(r)$ also nur noch vom Korrelationskoeffizienten der Grundgesamtheit ρ ab. Die Hypothese des Tests auf Unabhängigkeit lautet $H_0: \rho = 0$. (Allgemeiner kann $H_0: \rho = \rho_0$ formuliert werden.)

Dadurch vereinfacht sich die Dichtefunktion

$$\text{zu } f(r) = \frac{n-2}{\pi} \cdot (1-r^2)^{\frac{1}{2}(n-4)} \cdot \int_0^1 \frac{\xi^{n-2}}{\sqrt{1-\xi^2}} d\xi$$

$$\text{oder } f(r) = \frac{1}{\sqrt{\pi}} \cdot \frac{\Gamma(\frac{1}{2}(n-1))}{\Gamma(\frac{1}{2}(n-\frac{1}{2}))} \cdot (1-r^2)^{\frac{1}{2}(n-4)} \quad 2)$$

Falls $\rho = 0$ ist, kann statt R aber auch die Prüfgröße $\sqrt{n-2} \cdot R / \sqrt{1-R^2}$ betrachtet werden, die einer $t(n-2)$ -Verteilung folgt.³⁾ Diese Aussage

-
- 1) Vgl. z. B. FISZ, M.: Wahrscheinlichkeitsrechnung und mathematische Statistik, 7. Auflage, Berlin 1973
 - 2) Diese Zusammenhänge sind z. B. nachzulesen bei LEHMANN, E. L.: Testing Statistical Hypotheses, New York, 1959, S. 208 f.
 - 3) Die Approximation der Verteilung von R durch eine $N(\rho; (1-\rho^2)/\sqrt{n})$ -Verteilung soll nur erwähnt werden, da sie erst ab $n \geq 500$ hinreichend genau ist.

wird für diese Arbeit benutzt, da es sich um einen Test auf Unabhängigkeit handelt.

Es sei noch darauf hingewiesen, daß Hypothesen der allgemeinen Form $H_0: \rho = \rho_0$ über die Prüfgröße $Z = \frac{1}{2} \cdot \ln \frac{1+R}{1-R}$ getestet werden können, die asymptotisch einer $N(\frac{1}{2} \ln \frac{1+\rho_0}{1-\rho_0}; \frac{1}{n-3})$ -Verteilung folgt.¹⁾

Diese Prüfgröße kann auch für $\rho_0 \neq 0$ eingesetzt werden im Gegensatz zu der t-verteilten Größe

$$\sqrt{n-2} \cdot R / \sqrt{1-R^2} \quad 2)$$

Eine wesentliche Voraussetzung für den soeben beschriebenen Unabhängigkeitstest ist die Annahme der zweidimensionalen Normalverteilung von (X,Y), diese ist aber oft in der Praxis nicht erfüllt - und auch schwer nachzuprüfen. Deshalb wird man in der empirischen Sozialforschung auch bei metrischen Daten häufig auf Verfahren für niedrigeres Skalenniveau, z. B. den Test des Spearman-Koeffizienten, ausweichen müssen. Daß dies nicht überhaupt häufiger geschieht, ist verwunderlich bei den strengen Voraussetzungen des Bravais/Pearson-Verfahrens.

Zusammenfassung:

Anzahl der Stichproben	eine
Meßniveau	metrisch
Untersuchungsvariablen	zwei, mit gemeinsamer $N(a, \Sigma)$ -Verteilung
Hypothese	$\rho=0$
Alternative	ein- oder zweiseitig

-
- 1) FISHER, R. A.: On the "probable error" of a coefficient of correlation deduced from a small sample. In: Metron 1 Nr. 4 (1921), S. 1ff.
 - 2) Die Ähnlichkeit der Prüfgröße mit der des Tests für den Rangkorrelationskoeffizienten ist offensichtlich.

Prüfgröße	$\sqrt{n-2} \cdot R / \sqrt{1-R^2}$
Stichprobenverteilung	$t(n-2)$

3.7 Zusammenfassende Übersicht

Die vorausgegangenen Abschnitte 3.2 bis 3.6 enthalten insgesamt 25 Testverfahren für verschiedene Skalenniveaus und verschiedene Problemtypen. Es war nicht möglich, alle diese Tests in das Programmsystem REP-TEST aufzunehmen, das ohnehin schon hart an der Grenze der Kapazität des zur Verfügung stehenden APPLE II operiert. Welche Verfahren auszusortieren waren, orientierte sich vor allem an der Frage, welche Tests für die Sozialwissenschaften am ehesten geeignet schienen und welche am häufigsten eingesetzt werden. Damit fielen die wenigen parametrischen Tests alle heraus (oder fast alle, wenn man den Binomialtest zu den parametrischen Verfahren zählt). Ein so interessantes Verfahren wie der Randomisierungstest mußte ausgeschieden werden wegen des zu großen Rechenaufwandes. Übrig geblieben ist jedoch ein immer noch vielseitiges Angebot, das teilweise immerhin die Möglichkeiten der SPSS-Prozedur NPAR TESTS übersteigt. So sind zusätzlich zu den dort vorgesehenen Verfahren die im letzten Abschnitt behandelten drei Tests auf Unabhängigkeit aufgenommen. Auf den weniger gebräuchlichen Moses-Test wurde stattdessen verzichtet.

Da eine Beschränkung mit Rücksicht auf den Umfang dieser Arbeit und vor allem mit Rücksicht auf die Rechnerkapazitäten notwendig war, konnten

interessante Bereiche für die Anwendung von Signifikanztests gar nicht berücksichtigt werden. Dies gilt z. B. für die Gruppe der "Ausreißertests" mit den interessanten Verfahren von David/Hartley/Pearson,¹⁾ Grubbs²⁾ oder Dixon.³⁾ Diese neueren Verfahren haben sich noch nicht so durchsetzen können, wie die zuvor besprochenen, die teilweise schon als klassische Methoden bezeichnet werden müssen. Dabei ist das Problem der Ausreißer bei Messungen von großer Bedeutung für die Einschätzung der Ergebnisse, die aus den erhobenen Daten gewonnen wurden. Sind Ausreißer, also stark abweichende Messungen, im Datenbestand enthalten, so können sie teilweise durch systematische Fehler (Meß- oder Datenerfassungsfehler) erklärt werden. Ansonsten stellt sich die Frage, ob der jeweilige Merkmalsträger überhaupt in die Stichprobe gehört, d. h. ob er zur definierten Grundgesamtheit gehört. Die Hypothese der Ausreißertests lautet nun: Die abweichenden Werte gehören zur Stichprobe. Wird die Hypothese zurückgewiesen, so werden die betreffenden Daten aus weiteren Auswertungen ausgeschlossen.

Ein weiterer bisher ausgesparter Komplex soll noch erwähnt werden, nämlich die Trendtests, worunter

-
- 1) DAVID, H. A./ HARTLEY, H. O./PEARSON, E. S.: The distribution of the ratio in a single normal sample, of range to standard deviation. In: Biometrika, Vol. 41 (1954), S. 482-493
 - 2) GRUBBS, F. E.: Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments. In: Technometrics, 15 (1973), S. 53-66; GRUBBS, F. E./BECK, G.: Extension of sample sizes and percentage point for significance tests of outlying observations. In: Technometrics, 14 (1972), S. 847-854
 - 3) DIXON, W. J.: Ratio involving extreme values. In: Annals of Mathematical Statistics, 22 (1951), S. 68-78

die Verfahren von Mann,¹⁾ Cox/Stuart²⁾ und Wallis/Moore³⁾ fallen. Die Hypothese lautet: Die Zeitreihe enthält keinen Trend (keinen Aufwärtstrend, keinen Abwärtstrend). Getestet wird gegen die Alternative, es bestehe ein Trend (ein Aufwärtstrend, ein Abwärtstrend).

Mit diesen kurzen Anmerkungen sollte ein kleiner Ausblick gegeben werden auf weitere Möglichkeiten, die aber in der vorliegenden Arbeit nicht berücksichtigt werden konnten. Die Zahl parametrischer und nicht-parametrischer Testverfahren ist mittlerweile so unüberschaubar groß geworden, daß eine Auswahl für die Darstellung unerlässlich war. Diese Auswahl erfolgte nach dem etwas konservativen Gesichtspunkt der Gebräuchlichkeit und somit der Anwendbarkeit der entwickelten Programme.

Die nachfolgende Übersicht stellt die behandelten Testverfahren dar, nach Problemtyp und Skalenniveau geordnet. Für nicht im Programmsystem enthaltene Tests wurde ein Hinweis aufgenommen.

-
- 1) MANN, H. B.: Nonparametric tests against trend. In: Econometrika, 13 (1945), S. 245-259
 - 2) Cox, D. R.: Planning of experiments, New York 1958
 - 3) WALLIS, W. A./MOORE, G. H.: A significance test for time series analysis. In: Journal of the American Statistical Association, 36 (1941), S. 401-409

Meßniveau Problemtyp		nominal	ordinal	metrisch
1) Anpassung an eine theoretische Verteilung		Binomialt. χ^2 -Anpassungstest	Kolmogoroff/ Smirnow-Einstichproben.	
2) Zufälligkeit einer Stichprobe		—	Iterations-test	
3) Unterschiede zw. unabhäng. St. a) in den Verteilungen	2 Stichproben	Fishtest χ^2 -Homogenitätst.	Kolmogoroff/ Smirnow-Zweistichproben. Wald/Wolfowitz-Test	
	mehr als 2 Stichp.	χ^2 -Homogenitätstest		
b) in den Mittelwerten	2 Stichproben	—	Mediantest Mann/Whitney-U-Test	t-Test ⁺ Randomisierungst. ⁺ einfakt. Varianz. ⁺
	mehr als 2 Stichp.	—	Mediantest Kruskal/Wallis-H-Test	einfaktorielle Varianzanalyse ⁺
c) in den Streuungen	2 Stichproben	—	—	Varianzquotiententest ⁺
	mehr als 2 Stichp.	—	—	Bartlett-Test ⁺
4) Unterschiede zw. verb. Meßreihen				
a) in den Verteilungen	2 Stichproben	McNemar		
	mehr als 2 Stichp.	Cochran		
b) in den Mittelwerten	2 Stichproben	—	Dixon/Mood-T. Wilcoxon-test	t-Test für verbundene Stichp. ⁺
	mehr als 2 Stichp.	—	Friedmantest	
5) Zusammenhang zw. 2 Merkmalen		χ^2 -Unabhängigk.-T.	Spearman	Bravais/Pearson

Die mit einem + versehenen Tests sind nicht im Programmsystem REP-Test enthalten.

Kapitel 4 Programmbeschreibung von REP-TEST

Das System REP-TEST enthält außer den schon besprochenen Teilen FILE INPUT, DATA INPUT und DATA BASE SYSTEM zwei Programme zur Durchführung von statistischen Testverfahren.

Das Programm TESTVERFAHREN bietet etwa den Standard der SPSS-Prozedur NPAR TESTS plus einiger Erweiterungen. Allerdings ist REP-TEST als interaktives System angelegt, d. h. für die ständige Kommunikation mit dem Rechner gedacht. Deshalb erscheinen alle Ergebnisse als Bildschirmanzeige, ein Druckerprotokoll kann allerdings ohne Schwierigkeiten durch Anfertigen einer Hard-Copy des Bildschirminhalts erstellt werden.¹⁾ Die vorrangige Benutzung des interaktiven Betriebs und die Beschränkung auf Bildschirmanzeige hat Auswirkungen auf die Möglichkeiten des Programms. Es können nämlich für einzelne Verfahren nur eine begrenzte Anzahl von Stichproben bzw. nur eine begrenzte Anzahl von Kategorien der Variablen zugelassen werden.

Das Programm TESTS MIT HILFEN enthält nur 16 von den 19 im System vorhandenen Tests, um eindeutige Entscheidungen zu ermöglichen. Dafür bietet es Hilfestellung bei der Auswahl eines geeigneten Verfahrens.

1) Zur Verfügung stand ein Drucker NEC PC-8023B-C. Die Druckeranweisungen können zusätzlich ins Programm aufgenommen oder im Direktmodus bei Bedarf eingegeben werden.

4.1 Hauptmenu

Zu Beginn des Programms werden aus DATA-Statements die Namen der möglicherweise zu ladenden Programme und die Ladedauer in zuvor dimensionierte Arrays P% und D% eingelesen. Danach werden die Fehler-routinen initialisiert und es erfolgt der Sprung zum Menu, das in Zeile 50 beginnt. Es erscheint folgende Bildschirmanzeige:

```
***REP-TEST***  
  
NICHT-PARAMETRISCHE TESTVERFAHREN  
  
FILE INPUT.....1  
DATA INPUT.....2  
DATA BASE MANAGEMENT SYSTEM.....3  
TESTVERFAHREN.....4  
TESTS MIT ENTSCHEIDUNGSHILFEN...5  
ENDE DES PROGRAMMS.....6  
GETROFFENE AUSWAHL.....
```

Eingabefehler bei der Auswahl werden über die Fehler-routinen ab Zeile 60000 bzw. eine zusätzliche IF-Abfrage abgefangen. Die Zeile 100 verzweigt zu den verschiedenen Programmen des Systems. Bei den Optionen 1 bis 3 erfolgt zunächst der Sprung in das gemeinsame Unterprogramm, das die Lademeldung auf dem Bildschirm bewirkt (Zeile 30), anschließend wird das jeweilige Programm geladen und gleichzeitig gestartet. Bei den Optionen 4 und 5 ist vor den Ladeprozeß noch jeweils eine Sicherheitsabfrage (Zeilen 140 - 170 bzw. 190 - 230)

eingeschoben. Die Dauer des Ladevorgangs ist bei diesen beiden Programmen, die zudem leicht fälschlicherweise ausgewählt werden können, so lang, daß ein Irrtum vermieden werden soll. Die Entscheidung für eine dieser beiden Optionen muß deshalb nach einer kurzen Erläuterung über die beiden Möglichkeiten noch einmal bekräftigt werden, dann erst erfolgt die Lademeldung und das Laden und Starten.

Bei zunächst falscher Auswahl kann ins Menu zurückgegangen werden.

Die Beendigung des Programms erfolgt nur über das Menu mit Option 6.

4.2 Programm "TESTVERFAHREN" - Direkte Auswahl der Tests

Das Programm, das 19 Testverfahren für unterschiedliche Problemtypen und Skalenniveaus enthält, ist gedacht für den geübten Anwender statistischer Methoden, der ohne weiteres für eine gegebene Situation das geeignete Testverfahren auswählen kann. Es erschien nicht sinnvoll, den Aufruf der einzelnen Testverfahren über eine GOSUB-Struktur zu gestalten. Die einzelnen Test greifen ihrerseits auf common subroutines, also allen Verfahren gemeinsame Unterprogramme, zurück und allzu viele geschachtelte GOSUB ... RETURN-Statements können zu Fehlern des Interpreters führen. Zudem ist es günstiger, nach Durchführung eines Tests unmittelbar zum Menu und nicht erst zum Menuverteiler zurückzukehren, was durch ON ... GOTO und ein weite-

res GOTO zum Ende des Verfahrens statt mit ON ... GOSUB und RETURN erreicht werden konnte. Aus den in Abschnitt 2.1 genannten programmier-technischen Gründen beginnt das Programm mit den gemeinsamen Unterprogrammen, ab Zeile 1000 folgt das Menu für die Auswahl der Tests.

Die einzelnen Verfahren finden sich ab Zeile 11000, jeder Test beginnt mit einer runden Tausenderzahl als Zeilennummer (11000, 12000, 13000, usw.).

Den Abschluß bilden ab Zeile 61000 die Fehlerrou-tinen, in denen grobe Bedienungsfelder der Anwen-der aufgefangen werden.

Das Programm beansprucht etwas über 19 Kbyte Speicherplatz und ungefähr 11 K für die einfachen und dimensionierten Variablen. Zusammen mit dem vom System für Schleifen, Klammern, GOSUB-Befehle usw. beanspruchten Platz wird somit fast der ge-samte zur Verfügung stehende Platz ausgeschöpft.

4.2.1 Initialisierung und Menu

Die Zeilen 10 bis 30 dienen der Initialisierung des Programms. Es werden zunächst die benötig-ten Arrays dimensioniert, danach häufig benutzte Funktionen definiert, nämlich eine Rundungsfunk-tion R und die Verteilungsfunktion der Standard-normalverteilung STN. Dabei handelt es sich um die DERENZO-Approximation der Normalverteilung.¹⁾

1) DERENZO, S. E.: Approximations for Hand Calculators Using Small Integer Coefficients. In: Mathematics of Computation, 1977/31/137, S. 217

Es ist allgemein üblich, statt der exakten Berechnung von Verteilungen Approximationen durchzuführen, da sie wesentlich weniger Zeit beanspruchen und oft sehr gute Ergebnisse bringen. So beträgt der durchschnittliche Fehler der DERENZO-Approximation auf dem APPLE II $5 \cdot 10^{-5}$ bei einer Durchführungszeit von $5,6 \cdot 10^{-2}$ Sekunden. Vergleicht man diese Zeit mit den 9 Sekunden, die die exakte Berechnung benötigt, leuchtet der Vorteil von Approximationen unmittelbar ein.

Die DERENZO-Approximation arbeitet nach der Formel:

$$\Phi(Z) \sim \text{EXP}(-((83 \cdot Z + 351) \cdot Z + 562) \cdot Z / (703 + 165 \cdot Z)) / 2.$$

Für die Programmierung wurde Z noch durch ABS(Z) ersetzt.

Dieser Näherungsformel wurde der Vorzug gegeben vor anderen, z. B. der von ZELEN/SEVERO,¹⁾ HART²⁾ oder ABRAMOWITZ/STEGUN,³⁾ wegen ihrer Einfachheit bei relativ hoher Güte. Wiederum fiel auf, daß für Großrechner geeignete Formeln teilweise auf dem APPLE II relativ schlechte Approximationsergebnisse brachten.

Es folgt im Programm die Initialisierung der Fehlerroutrinen und die Definition einiger Konstanten.

-
- 1) siehe dazu JOHNSON, N. L./KOTZ, S.: Continuous Multivariate Distributions, 2 Bde, New York, 1970, Formel 26, S. 55
 - 3) NORUSIS, M. J.: SPSS - Statistical Algorithms, Chicago, 1979, S. 152
 - 2) ebd., Formel 28', S. 56

Anschließend erscheint auf dem Bildschirm die Aufforderung, die Datendiskette einzulegen. Ist dies geschehen, wird der VARCATALOG eingelesen und es erfolgt der Sprung zum Menu (Zeilen 1000 bis 1060) mit den Testverfahren. Vor Beginn werden zunächst mit der Funktion FRE (0) alle alten Strings gelöscht, sodann erscheint nachfolgender Bildschirm Ausdruck:

REP-TEST

BINOMIAL.....1	CHI QU.-UNABH....11
CHI QU.-1 ST.....2	SPEARMAN.....12
KOLMOG./SM.-1 ST.3	BRAVAIS/PEARSON..13
CHI QU.-MEHR ST..4	MEDIAN.....14
FISHER.....5	MANN/WHITNEY....15
KOLM./SM.-2 ST...6	KRUSKAL/WALLIS...16
WALD/WOLFOWITZ...7	DIXON/MOOD.....17
MENEMAR.....8	WILCOXON.....18
COCHRAN.....9	FRIEDMAN.....19
RUNS.....10	REP-TEST MENU...20
AUSWAHL.....	

Fehler in der Eingabe werden über die Fehler-routinen ab 61000 bzw. eine zusätzliche IF-Abfrage abgefangen, soweit dies möglich ist. Falls die Option 20 gewählt wird, erscheint die Aufforderung, die Systemdiskette einzulegen, damit die Rückkehr zum Hauptmenu erfolgen kann (Zeilen 1070 -

1080). Nur so ist eine normale Beendigung des Programms möglich.¹⁾

4.2.2 Unterprogramme

In Zeile 100 beginnen die sogenannten common subroutines, die von allen bzw. mehreren Programmen aufgerufen werden. Zunächst steht in Zeile 100 bis 170 der X/Y-Sorter von VELLEMAN/HOAGLIN,²⁾ der von SCHNELL leicht verbessert wurde. Die subroutine benötigt N, X(I) und Y(I) als Eingänge und liefert ein der Größe nach aufwärts sortiertes Array X und ein entsprechend umgestelltes Array Y als Ausgänge.

Auf den Superquicksort von KOLBE mußte wegen Mangels an Speicherplatz verzichtet werden. Auch wenn nur ein X-Array zu sortieren ist, wird auf den X/Y-Sorter zurückgegriffen, der etwas langsamer ist (maximal 31 Sekunden bei 200 Fällen). Auf ihn kann aber keinesfalls verzichtet werden.

Die Zeilen 180 bis 200 enthalten die weiterentwickelte Ranking-Routine von SCHNELL, die zudem noch die Anzahl der ties und die Korrekturfaktoren beim Auftreten von Verbundwerten für einige Tests bestimmt. Das Unterprogramm benötigt als Eingänge N und das Array X, das zuvor schon sortiert worden sein muß. Ausgang ist ein Array X,

1) Natürlich kann auch ein Bedienungsfehler zur Beendigung des Programms führen, ohne Rückkehr zum Hauptmenu. Es erscheint eine FATAL ERROR-Meldung. Falls Fehler bei der Dateneingabe/-ausgabe auf Diskette auftreten, wird das Programm nach einer entsprechenden Fehlermeldung neu gestartet

2) VELLEMAN/HOAGLIN (1981), S. 304

das statt der ursprünglichen, sortierten Werte nun Rangplätze von 1 bis N, bzw. bei Verbundwerten mittlere Rangplätze, enthält.

Es folgen nun die Approximationen der t-Verteilung und der χ^2 -Verteilung. Schon bei der Definition der Funktion STN wurde erläutert, daß Approximationen für komplizierte Funktionen unumgänglich sind, ganz besonders bei Kleinrechnern, wegen der oft langen Ausführungszeiten für die Berechnung der exakten Werte.

Die Zeilen 210 bis 250 enthalten die Berechnung der Verteilungsfunktion der t-Verteilung. Eingänge sind der Wert t, für den die Verteilungsfunktion berechnet wird, und die Anzahl der Freiheitsgrade V. Für $V \leq 30$ wird die exakte Formel nach ABRAMOWITZ/STEGUN¹⁾ verwendet, die mit $X=ABS(t)$ auf folgende Weise ein zweiseitiges Signifikanzniveau $S=1-A(V)$ bestimmt:

$$A(1) = 2 \cdot y / \pi$$

$$A(V) = \frac{2}{\pi} \cdot \{ y + \sin y \cdot \left[\cos y + \sum_{i=1}^{k-1} \left(\prod_{j=1}^i \frac{2j}{2j+1} \right) \cdot \cos^{2i+1} y \right] \},$$

falls $V=2k+1$, $k \geq 1$

$$A(V) = \sin y \cdot \left[1 + \sum_{i=1}^{k-1} \left(\prod_{j=1}^i \frac{2j-1}{2j} \right) \cdot \cos^{2i} y \right],$$

falls $V=2k$, $k \geq 1$.

Dabei ist $y = \arctan y'$ mit $y' = X/\sqrt{V}$ und $\sin y = y' \cdot (1+y'^2)^{-1/2}$, $\cos y = (1+y'^2)^{-1/2}$.

1) NORUSIS (1979), S. 153

Der maximale Fehler dieser Formel liegt bei 10^{-5} , die Ausführungszeit ist nur unwesentlich länger als bei den meisten Näherungsformeln, deren viele geschachtelte Klammern sehr viel Zeit kosten.

Für $V > 30$ lohnt sich allerdings der Übergang zur Approximation von ZELEN/SEVERO,¹⁾ die einen mittleren Fehler unter 10^{-4} hat. Die Näherung beruht auf der Konvergenz der t-Verteilung gegen die Standardnormalverteilung. Das zweiseitige Signifikanzniveau S bestimmt sich aus:

$$S = 2 \cdot \Phi(t(1 - \frac{1}{4} V^{-1}) (1 + \frac{1}{2} t^2 V^{-1})^{-1/2});$$

dies ist der Ausgang der subroutine.

Die Zeilen 260 bis 280 enthalten die nach den Formeln von BORCHERS/POOLE²⁾ gerechnete Approximation der χ^2 -Verteilung³⁾ für $V \leq 35$ und $10^{-7} \leq \chi^2 \leq 90$. Eingänge des Programms sind V und CH , Ausgang ist das Signifikanzniveau S . Die Formeln lauten:

$$S = 1 - CH^{(V+1)/2} \cdot e^{-CH/2} \cdot \left(\prod_{i=1}^k (2i+1) \right)^{-1} \cdot (2/CH/\pi)^{1/2} \cdot z$$

$$S = 1 - CH^{V/2} \cdot e^{-CH/2} \cdot \left(\prod_{i=1}^k 2i \right)^{-1} \cdot z, \text{ falls } V = 2k+1, k \geq 1$$

$$S = 1 - CH^{V/2} \cdot e^{-CH/2} \cdot \left(\prod_{i=1}^k 2i \right)^{-1} \cdot z, \text{ falls } V = 2k, k \geq 1.$$

$$\text{Dabei ist } z = 1 + \sum_{m=1}^{\infty} CH^m \cdot \left(\prod_{i=1}^m (V+2i) \right)^{-1}.$$

Werte von V und CH außerhalb des genannten Bereichs führen zu $S=999$.

1) JOHNSON/KOTZ (1970), Formel 34, S. 113

2) BORCHERS, L./POOLE, M.: Some Common BASIC Programs, Berkely, 1979, S. 130 ff.

3) Die Formeln entsprechen denen des Algorithmus AS 32 aus Bhattacharjee, G. P.: Algorithm AS 32: The Incomplete Gamma Integral. In: Applied Statistics 1970/19, S. 285-287

Der mittlere Fehler dieser Näherung liegt bei 10^{-5} , der maximale beträgt immer noch nur 10^{-4} . Andere gute oder bessere Approximationen mußten wegen zu aufwendiger Programmierung und zu langer Ausführungszeit verworfen werden.¹⁾

Die nun folgenden Zeilen 290 und 300 berechnen das Signifikanzniveau unter Binomialverteilung. Eingänge sind P (der Bernoulli-Parameter), H9 (der Stichprobenumfang N) und K (die Stichprobenrealisation). Die notwendigen Binomialkoeffizienten werden aus dem nachfolgenden Unterprogramm abgerufen. Der Rechenaufwand konnte auf ein Minimum reduziert werden durch Festlegen der Startgröße $S_0 = \pi^0 (1-\pi)^N$ und $S_i = S_{i-1} + S_{i-1}\pi / (1-\pi)$. Ausgang des Programms ist S, das einseitige Signifikanzniveau.

Die Zeilen 310 bis 330 sind zur Berechnung der Binomialkoeffizienten nötig. Die subroutine wird nicht nur von der vorhergehenden abgerufen, sondern auch unabhängig von dieser. Eingänge sind H9 (für N) und H8 (für K). Dabei kann H8 sich im Laufe des Programms verändern, da es möglicherweise durch H9-H8 ersetzt wird, damit die Rechnung vereinfacht wird. Durch diesen "Kürzungsverfahren" ist es möglich, Binomialkoeffizienten bis zu $\binom{50}{24}$ zu berechnen. Ausgang des Programms ist H1.²⁾

-
- 1) Vgl. z. B. TASSEL (1981), S. 158 ff. oder LING, R. F.: A Study of the Accuracy of Some Approximations for the t, χ^2 , and F Tail Probabilities. In: Journal of the American Statistical Association, 1978/73, S. 274-283
 - 2) Es war nötig, soviel wie möglich mit einer Reihe von Hilfsvariablen H1-H9, H1%-H9% zu arbeiten, die ständig wechselnde Bedeutungen haben, um Platz im numerischen Speicher zu sparen. Die dadurch entstehende Unübersichtlichkeit mußte in Kauf genommen werden.

Dies waren die subroutines für Umgestaltungen der Datenarrays und für Berechnungen, die für die Tests immer wieder benötigt werden. Es folgen nun Prozeduren wie Variablenladen und Variablenidentifikation, häufig wiederkehrende Bildschirmausdrucke u. a..

Die Zeilen 400 bis 510 enthalten alle Unterprogramme, die nötig sind zur Eingabe der gewünschten Variablen, zu ihrer Identifikation im Variablenkatalog und dem Ladevorgang. Aus dem Hauptprogramm muß die Anzahl der für die Prozedur benötigten Variablen übergeben werden. Dies geschieht meist durch das Programm, außer bei den Verfahren für mehr als zwei verbundene Stichproben, die für VC% eine Eingabe verlangen. Die Eingabe der gewünschten Variablen erfolgt durch Eingabe des Strings A\$, der jeweils den Variablennamen enthält. Anhand dieses Namens wird die Liste aller Variablennamen des VARCATALOG - sie befindet sich im Array V\$ - durchsucht.

Falls der genannte Name nicht existiert, erfolgt eine Fehlermeldung; ansonsten wird die entsprechende Array-Position einem "Prozedurkatalog" in dem Integerarray VM%(I) (mit I=1, ..., VC%) zugeordnet. Die Identifikationsnummern (aus dem V\$-Array) der bereits in der CPU vorhandenen Variablen sind in einem "CPU-Katalog" gespeichert. Dies ist das Integerarray VC%, das 5 Positionen hat. Wenn die für eine Prozedur benötigten Variablen bereits in der CPU vorhanden sind, kann zum Hauptprogramm zurückgekehrt werden. Ansonsten wird das Array VC% auf nicht benötigte Variablen unter-

sucht und diese werden gegebenenfalls durch die neu zu ladenden ersetzt. Diese Entscheidungen werden durch Vergleich der beiden Arrays VM% und VC% erreicht.

Die Zeilen 400 und 410 bewirken den eigentlichen Ladevorgang und eine entsprechende Meldung auf dem Bildschirm.

Die Zeile 420 fordert die Eingabe der gewünschten Variablen in einer Schleife bis zur maximalen Anzahl VC%. Danach erfolgt der Sprung in die Identifikationssubroutine (Zeilen 440 bis 460). Nach der Rückkehr wird die gefundene Identifikationsnummer dem Prozedurkatalog VM% an entsprechende Position zugewiesen und zur Sicherheit der Name der Variablen noch einmal angezeigt. In einer weiteren Schleife wird für alle gewünschten Variablen die subroutine zum Suchen der Variablen in der CPU (Zeilen 470 bis 510) angesprungen, die gegebenenfalls auch den Ladeprozeß auslöst. Nach Ende dieser Schleife stehen alle benötigten Variablen zur Verfügung und es erfolgt die Rückkehr zum Hauptprogramm.

Die Zeile 430 wird nur abgearbeitet, wenn nicht-existierende Variablennamen eingegeben werden. Zeile 440 fängt Eingabefehler ab und bewirkt die Rückkehr zum Menu.

Die Zeilen 450 und 460 enthalten den Identifikationsprozeß.

In Zeile 470 wird überprüft, ob eine benötigte Variable in der CPU vorhanden ist; ist dies der Fall, kann zur aufrufenden Zeile 420 zurückgekehrt werden. Wenn nicht, erfolgt die Überprüfung, ob die

Anzahl der "CPU-Variablen" (NC%) schon maximal (d. h. gleich fünf) ist, d. h. ob noch zusätzliche Variablen geladen werden können. Wenn ja, folgt die Erhöhung des Zählers NC% und der Sprung in Zeile 51Ø.

Falls die Anzahl NC% bereits fünf beträgt, wird in den Zeilen 49Ø und 50Ø geprüft, welche CPU-Variable überschrieben werden kann.

In Zeile 51Ø wird sodann der CPU-Katalog aktualisiert und danach wird die Ladesubroutine 40Ø bis 41Ø angesprungen. Nach dem Ladevorgang erfolgt die Rückkehr zur aufrufenden Zeile 42Ø.

Dieses etwas komplizierte Verfahren ist notwendig, um unnötige Ladevorgänge zu vermeiden. Dies wird erreicht durch das Führen zweier Kataloge VM% und VC%, die nur Identifikationsnummern enthalten, so daß die Variablennamen nur einmal im Variablenkatalog V% gespeichert sein müssen.

Die nun folgenden sechs Zeilen 52Ø bis 57Ø bewirken das Aussortieren von missing values für unterschiedliche Situationen. In 52Ø und 53Ø wird das Array A(VM%(1),J) (für alle J=1, ..., NF%) einer Variablen auf missing values untersucht. Die gültigen Fälle werden durch C% gezählt und nur sie werden einem Array X(J) zugeordnet (wobei nun J=1, ..., C% ist).

Die Zeilen 54Ø bis 55Ø leisten ähnliches für zwei Variablen. Eingänge sind also die Arrays A(VM%(1),J) und A(VM%(2),J) (J=1, ..., NF%). Wenn mindestens eine der beiden Variablen für einen Fall den missing value annimmt, wird dieser Fall ausgeschlossen. Die

gültigen Fälle werden wieder durch C% gezählt und den Arrays X(J) und Y(J) (J=1, ..., C%) zugewiesen.

Die Zeilen 560 und 570 sind für den Fall vorgesehen, daß mehr als 2 Variablen eingelesen werden müssen. Es wird dann das Array A(VM%(I),J) (I=1, ..., VC%; J=1, ..., NF%) verändert, nämlich reduziert um die Fälle, in denen mindestens eine Variable den missing value annimmt. Die gültigen Fälle werden wieder durch C% gezählt, so daß die Ausgänge der subroutine C% und das veränderte Array A(VM%(I),J), (I=1, ..., VC%; J=1, ..., C%) sind.¹⁾

Die nachfolgenden Zeilen 600 bis 650 enthalten kurze Unterprogramme für Bildschirmanzeigen.

Zeile 600 wird zu Beginn jedes Testverfahrens aufgerufen und bewirkt eine Überschrift mit dem Namen des jeweiligen Tests.

Zeile 610 ruft eine Bildschirmanzeige hervor mit der Dauer der Rechnung.

Zeile 620 liefert für die Prozeduren mit einer Variablen den Variablennamen als Überschrift über das Testergebnis.

Zeile 630 liefert eine ähnliche Überschrift für die Prozeduren, die zwei Variablen benötigen.

Zeile 640 schließlich leistet das gleiche für die Tests mit mehr als zwei Variablen.

Durch Zeile 650 erfolgt die Bildschirmanzeige, die alle Testergebnisse abschließt, nämlich

1) Dieser Veränderung des Eingabearrays wurde bei den entsprechenden Prozeduren Cochran-Q-Test und Friedman-Test Rechnung getragen. Darauf soll im Abschnitt 4.2.3 ausführlich Bezug genommen werden.

die Anzahl der gültigen und fehlenden Fälle.

Die Zeilen 660 und 670 dienen zur Initialisierung benötigter Zähl- und Hilfsvariablen, nämlich der Arrays S,T% und N% und der einfachen Variablen H1 bis H9 und H1% bis H9%.

In Zeile 700 befindet sich eine Warteschleife mit bedingtem Abbruch. Dieses Unterprogramm wird benötigt, um Fehlermeldungen und Testergebnisse auf dem Bildschirm zu halten, bis der Benutzer im Programm fortfahren will.

Die Zeile 710 bewirkt eine Fehlermeldung, falls keine gültigen Fälle mehr vorhanden sind, d. h. falls C%=0 ist. In jedem Testverfahren wird nach dem Aufbau der korrigierten Datenarrays (Zeilen 520 bis 570) überprüft, ob C%=0 ist, und gegebenenfalls zur Zeile 710 verzweigt. Anschließend erfolgt die Rückkehr ins Menu.

In Zeile 720 erfolgt die Überprüfung ob Variablen 0/1-codiert sind, wenn der jeweilige Test dies verlangt. Eingang ist das Array Y, das unverändert bleibt. Falls ein Wert ungleich Null oder ungleich Eins entdeckt wird, erfolgt in der nachfolgenden Zeile 730 eine Fehlermeldung, danach die Rückkehr zum Menu.

Zeile 740 wird angesprungen, wenn in einem Testverfahren eine Standardabweichung von Null festgestellt wird. Dann erfolgt eine entsprechende Fehlermeldung mit anschließender Rückkehr zum Menu.

Die nun folgenden Unterprogramme sind Teile von Testverfahren, die an mehreren Stellen des Hauptprogramms benötigt werden.

Die Zeilen 750 und 760 bewirken eine Bildschirmanzeige. Für den Mann/Whitney-U-Test und den Kruskal/Wallis-H-Test wird eine Ergebnistabelle erzeugt, die die Häufigkeiten in den verschiedenen Gruppen und die mittleren Ränge der Gruppen enthält. Das Unterprogramm enthält die Anweisungen zum Ausdruck dieser Tabelle. Eingänge sind die Häufigkeiten $T(I)$, die Rangsummen der Gruppen $S(I)$ und die Anzahl der Gruppen $H1\%$.

Die Zeilen 770 bis 790 enthalten die Berechnungen für das Signifikanzniveau von Fisher's exaktem Test. Diese subroutine wird sowohl von der Prozedur "Fisher-Test" als auch vom Verfahren "Median-Test" aufgerufen. Eingänge sind die Häufigkeiten einer Vierfelder-Tabelle $N(I,J)$ und die Anzahl der Fälle $C\%$, Ausgang ist "Fisher's exakte Wahrscheinlichkeit", das einseitige Signifikanzniveau S . Außerdem wird dieses Unterprogramm vom χ^2 -Homogenitätstest benötigt. In dieser Prozedur wird für Vierfeldertabellen bei Stichprobenumfängen bis 30 ebenfalls ein Fisher-Test durchgeführt.

Die Zeilen 800 bis 820 enthalten die Anweisungen zur Bestimmung des Median. Dieses Unterprogramm wird vom Median-Test aufgerufen und vom Run-Test, falls dort am Median dichotomisiert werden soll.

Außer der Bestimmung des Medians für die Variable X in der Zeile 800 wird X noch umcodiert in eine 0/1-Variable, wobei $X=0$ gesetzt wird, falls X kleiner oder gleich dem Median ist. Dies wird für den Mediantest benötigt. Eingänge des Unterprogramms, das seinerseits die Sort-subroutine (100-170) aufruft, sind C% und das Array X, das sortiert und umcodiert aus der Routine hervorgeht.

Zeile 830 enthält eine einfache Auszählung von gemeinsamen Häufigkeiten zweier Variablen und die Bestimmung der Randverteilungen. Eingänge sind C% und die Arrays X und Y und zudem das Minimum der Variablen Y, nämlich H1%. Ausgänge sind die bivariaten Häufigkeiten N%(I,J) und die univariaten Häufigkeiten T%(I) und S(J). Dabei beginnt der Wertebereich von J nicht beim Minimum von Y, sondern bei Null. Außerdem wird vor der Rückkehr noch die Subroutine aus Zeile 630 aufgerufen für eine Bildschirmanzeige, nämlich die Überschrift des Testergebnisses bei einem Verfahren mit zwei Variablen. Das Unterprogramm wird von den Verfahren "Mediantest" und "Fisher-Test" aufgerufen.

Die Zeile 840 bewirkt den Ausdruck einer kx2-Häufigkeitstabelle, wie sie vom Mediantest und dem Fisher-Test benötigt wird. Eingänge der subroutine sind N%(J,I), T%(J) und S(I) ($J=0,1$; $I=0, \dots, H3\%$), C% und H3%. Dabei ist $H3\%+1$ die Anzahl aller Ausprägungen einer Gruppierungsvariablen.

Die Zeile 850 enthält die Überschrift zur eben genannten Tabelle (Zeile 840) für den Mediantest.

Die nun folgenden zwei Unterprogramme werden ausschließlich von den beiden Kolmogoroff/Smirnow-Testverfahren benötigt.

Die Zeilen 860 bis 950 enthalten die Bestimmung der Prüfgröße und die aufwendige Berechnung des Signifikanzniveaus.

Eingänge sind die beiden Arrays X und Y, die aber nicht mit den ursprünglich eingelesenen Datenarrays identisch sind.

Für den 1-Stichprobentest enthält Y(J) die Werte der empirischen Verteilungsfunktion für alle verschiedenen Werte aus der Stichprobe und X(J) die entsprechenden Werte der theoretischen Verteilungsfunktion. Da möglicherweise Verbundwerte aufgetreten sind, gibt N die Anzahl der Kategorien an, also $N \leq C\%$. A% enthält die Option, die aus dem Untermenü dieses Tests gewählt wurde, über die gewünschte theoretische Verteilung. V ist hier gleich C%. Für den 2-Stichprobentest sind X(J) und Y(J) die Werte der beiden kumulierten Verteilungsfunktionen, und zwar enthält Y(J) die Funktion aus der größeren der beiden Stichproben; in diesem Fall wird A% nicht benötigt und vorher im Hauptprogramm gleich drei gesetzt. N ist gleich C% und V gleich dem Produkt der beiden Stichprobenumfänge, dividiert durch ihre Summe, nämlich C%.

Das Programm bestimmt die größte (positive) Differenz (H3) und die kleinste (d. h. die "größte negative") Differenz (H4), außerdem die absolut größte Abweichung der beiden Funktionen (H5), die

Grundlage für die Prüfgröße Z ist. Mit Hilfe der Approximationsformel von SMIRNOW (1948) wird ein zweiseitiges Signifikanzniveau S berechnet.¹⁾

Die Zeilen 960 bis 980 enthalten die Anweisungen für die Bildschirmanzeige der Testergebnisse der beiden Kolmogoroff/Smirnow-Tests. Dabei wird berücksichtigt, ob tatsächlich positive und negative Abweichungen aufgetreten sind. Eingänge dieser subroutine sind die Ausgänge der vorherigen, nämlich $H3$, $H4$, $H5$, Z und S .

Damit ist der Bereich der common subroutines abgeschlossen. Auf eine weitere "Zergliederung" des Programms wurde verzichtet zugunsten einer besseren Lesbarkeit und zugunsten der Rechenzeit. Da jeder Sprung in ein Unterprogramm Zeit kostet, hingegen das Vermeiden von Unterprogrammen zu "aufgeblähten" Programmen führt, muß immer wieder ein Kompromiß gefunden werden zwischen der Reduzierung des Platzbedarfs, der Herabsetzung der Rechenzeit und der Erhöhung der Übersichtlichkeit. Zum letzten Punkt soll die nachfolgende Übersicht etwas beitragen; sie soll die Orientierung in diesem Teil des Programms erleichtern. Zu jeder subroutine sind die Zeilennummern, eine kurze Beschreibung ihrer Aufgabe und ihre Ein- und Ausgänge aufgeführt.

1)

$$\text{Es ist } \lim_{n \rightarrow \infty} P(Z \leq \lambda) = 1 - 2 \cdot \sum_{k=1}^{\infty} (-1)^{k-1} \cdot e^{-2k^2 \cdot \lambda^2},$$

wofür von Smirnow eine Näherung entwickelt wurde.
Zum Beweis vgl. z.B. FISZ (1976).

Übersicht über die "common subroutines":

Zeile	Aufgabe	Input	Output
100-170	X-Array sortieren, Y-Array mitführen	N,X(J), Y(J),	X(J), X(J)
180-200	Vergeben von Rang- plätzen an sortier- tes X-Array	N,X(J)	X(J)
210-250	Signifikanz eines t-Wertes, zweiseitig	V,T	S
260-280	Signifikanz eines χ^2 -Wertes	V,CH	S
290-300	Signifikanz bei Binominalverteilung	P,H9, K	S
310-330	Berechnung von Bino- mialkoeffizienten	H9,H8	H1
400-410	Lesen eines Variablen- files	VC%(I%)	A(I%,J)
420	Eingabe der Va- riablenennamen	A\$	VM%(I%)
430	Wiederholung der Ein- gabe eines Variablen- namens (bei Fehlern)	A\$	-

Zeile	Aufgabe	Input	Output
44Ø-46Ø	Variablenidentifikation	AØ	I%
47Ø-51Ø	Suchen einer Variablen in der CPU und Laden	VM%(J)	VC%(I%)
52Ø-53Ø	Aussortieren von missing values bei 2 Variablen, Aufbau der X- und Y-Arrays, Bestimmung der Anzahl gültiger Fälle	A(VM%(1),J)	X(J) A(VM%(2),J) Y(J),C%
54Ø-55Ø	Aussortieren von m. v. für 1 Variable, Aufbau des X-Arrays, Bestimmung der Anzahl gültiger Fälle	A(VM%(1),J)	X(J),C%
56Ø-57Ø	Aussortieren von m. v. für mehr als 2 Variable, Neuaufbau des A-Arrays, Bestimmung der Anzahl gültiger Fälle	A(VM%(I),J)	(A(VM%(I),J) C%
60Ø	Bildschirmanzeige für Namen der Tests	AØ	-
61Ø	Bildschirmanzeige: Dauer der Rechnung	DU%	-

Zeile	Aufgabe	Input	Output
62Ø	Bildschirmanzeige: Variablenname als Überschrift des Testergebnisses für 1 Variable	VM%(1)	-
63Ø	Bildschirmanzeige: Variablennamen als Überschrift des Test- ergebnisses für 2 Va- riablen	VM%(1) , VM%(2)	-
64Ø	Bildschirmanzeige: Variablennamen als Überschrift des Testergebnisses für mehr als 2 Variablen	VM%(I)	-
65Ø	Bildschirmanzeige: Anzahl gültiger Fälle und missing values	C%	-
66Ø	Initialisierung der Arrays S,T% und N%	-	-
67Ø	Initialisierung der Hilfsvariablen H1 bis H9 und H1% bis H9%	-	-
7ØØ	Warteschleife mit be- dingtem Abbruch	-	-

Zeile	Aufgabe	Input	Output
71Ø	Fehlermeldung: keine gültigen Fälle, Abbruch	-	-
72Ø	Überprüfung von Ø/1-Codierungen	Y(J)	-
73Ø	Fehlermeldung: keine Ø/1-Co- dierung. Abbruch	-	-
74Ø	Fehlermeldung: Standardab- weichung ist Null, Abbruch	-	-
75Ø-76Ø	Bildschirmanzeige: Tabelle für Mann/ Whitney und Kruskal/ Wallis	H1%, T%(I) S(I)	-
77Ø-79Ø	Berechnung von Fisher's exakter Wahrscheinlich- keit	C%, N%(I,J)	s
80Ø-82Ø	Bestimmung des Medians des X-Arrays und Ø/1- Umcodierung	C%, X(J)	X(J)
83Ø	Bivariate Häufigkeits- auszählung mit Randverteilungen	H1%, X(J), Y(J)	N%(X,Y) T%(X), S(Y)

Zeile	Aufgabe	Input	Output
84Ø	Bildschirmanzeige: kx2-Tabelle für Me- dian und Fisher- Test	H3%, H1%, N%(J,I), T%(J), S(I)	-
85Ø	Bildschirmanzeige: Überschrift der kx2- Tabelle aus 93Ø für Mediantest	H9	-
86Ø-95Ø	Bestimmung der max. Differenzen für K-S- Tests und Berechnung des Signifikanznive- aus	N,X(J), Y(J),A%, V	H3,H4, H5,Z, S
96Ø-98Ø	Bildschirmanzeige: Ergebnis der K-S- Tests	H3,H4,H5, Z,S	-

4.2.3 Testverfahren

Die folgenden Abschnitte beschreiben die Programmzeilen 11000 bis 29050, die 19 Testverfahren enthalten. Die Beschreibung dieser Programnteile erfolgt nach ihrer Reihenfolge im Programm und nicht nach Problemtypen geordnet wie die Beschreibung der Testverfahren im Kapitel 3. Die Testverfahren beginnen in den Zeilen 11000, 12000 usw., in Tausenderschritten, um die Übersichtlichkeit zu erhöhen. Die erste Anweisung ist stets die Zuweisung des Namens des jeweiligen Verfahrens an die Variable A\$, so daß leicht identifiziert werden kann, welcher Test in den nachfolgenden Zeilen gerechnet wird. Anschließend erfolgt der Sprung in das Unterprogramm 600 für die Bildschirmanzeige des Namens. Die Anzahl der benötigten Variablen (VC%) wird je nach Verfahren auf 1 oder 2 gesetzt bzw. vom Benutzer wird die Eingabe eines Wertes verlangt. Danach wird in die Zeile 420 gesprungen, zur Variablenidentifikation und zum Ladevorgang. Anschließend erfolgt die Überprüfung der Eingabe auf missing values und der Aufbau der endgültigen Datenarrays in den subroutines 520, 540 oder 560. Bei einigen Verfahren erscheint vorher noch ein Hinweis, wenn bestimmte Codierungen der Variablen vorausgesetzt werden.

Falls die Anzahl der gültigen Fälle Null ist, erfolgt der Sprung zur Fehlermeldung in Zeile 710, die auch den Abbruch der Rechnung verursacht.

Danach erfolgt gegebenenfalls eine Überprüfung der korrekten Codierung (Zeile 720).

Dieser Ablauf ist bei fast allen Programmen gleich oder ähnlich, die entsprechenden Anweisungen werden deshalb in den nachfolgenden Beschreibungen nicht mehr erläutert.

4.2.3.1 Binomialtest

Dieses Verfahren beansprucht die Programmzeilen 11000 bis 11060. Zeile 11000 enthält die zuvor beschriebene "Initialisierung". In Zeile 11010 wird nach der Überprüfung der Codierung die Eingabe des Parameters π angefordert. Wenn diese unterbleibt, wird die Voreinstellung $\pi = 0,5$ für die Berechnungen verwendet. Zeile 11020 enthält eine häufig benutzte Folge von Anweisungen, durch die eine Eingabe außerhalb des zulässigen Wertebereichs als Fehler erkannt wird und erneute Eingabe gefordert wird.¹⁾

Zeile 11030 beginnt mit den Statements für die Bildschirmanzeige über die Dauer der Rechnung (Zeile 610), danach erfolgt in einer Schleife das Auszählen der Häufigkeiten eines "Erfolges", hier durch den Code "1" dargestellt. Für $C \leq 25$ (C =Anzahl gültiger Fälle) wird die exakte Verteilung aus dem Unterprogramm für die Binomialverteilung (ab Zeile 290) benutzt, danach erfolgt die Bildschirmanzeige (11050).

Für größere Fallzahlen wird die Approximation der Binomialverteilung durch die Normalverteilung verwendet. Zeile 11040 enthält die Berechnung der

1) Zu den verwendeten Maschinenunterprogrammen siehe Anhang

Prüfgröße einschließlich einer Stetigkeitskorrektur.

Berechnet wird in beiden Fällen ein einseitiges Signifikanzniveau.

Die Bildschirmanzeige (11050 bis 11060) enthält eine Tabelle mit den Häufigkeiten der beiden Ausprägungen, den relativen Häufigkeiten und den berechneten Wahrscheinlichkeiten. Zudem die Angabe des Bernoulli-Parameters π .

Den Abschluß bildet die Anzahl der gültigen und fehlenden Fälle (Zeile 650). Anschließend wird die bedingte Warteschleife in Zeile 700 initialisiert. Bis zum nächsten Tastendruck bleibt der Bildschirminhalt zu sehen, danach erfolgt die Rückkehr ins Menu.

4.2.3.2 χ^2 -Anpassungstest

Das Verfahren beansprucht 10 Programmzeilen (12000 bis 12090). Nach dem üblichen Beginn werden in Zeile 12010 zunächst die Hilfsvariablen und die Zählerarrays auf 0 gesetzt (Unterprogramme 660 und 670), danach werden in einer Schleife bis zum Beginn der Zeile 12030 Minimum (H1%) und Maximum (H2%) der Variablen bestimmt, anschließend die Anzahl der Kategorien (H3%). Ist H3% größer als 8, so erfolgt der Abbruch der Rechnung mit einer entsprechenden Bildschirmanzeige. Diese Beschränkung mußte im Hinblick auf die spätere Ergebnisanzeige aufgenommen werden. Danach erfolgt in 12040 die

Anfrage, ob auf Gleichverteilung geprüft werden soll, da dies erfahrungsgemäß¹⁾ die am häufigsten auftretende theoretische Verteilung beim χ^2 -Anpassungstest ist.

In Zeile 12050 werden zunächst die Häufigkeiten der einzelnen Kategorien T%(I) ausgezählt und, falls auf Gleichverteilung getestet werden soll, die erwarteten Häufigkeiten S(I) berechnet.

Im anderen Fall wird vom Benutzer durch die Zeile 12060 die Eingabe der erwarteten Häufigkeiten gefordert. Stimmt die Summe Y der eingegebenen Werte nicht mit der Anzahl gültiger Fälle überein, erfolgt eine entsprechende Anzeige und die Eingabe muß komplett wiederholt werden.²⁾ In Zeile 12070 beginnt die eigentliche Berechnung mit der entsprechenden Meldung auf dem Bildschirm (610). Es erfolgt die Bestimmung der Prüfgröße CH. Falls eine der erwarteten Häufigkeiten kleiner als fünf ist, wird das Verfahren abgebrochen und zum Menu zurückgekehrt. Nach einer Zusammenlegung von Kategorien kann der Test neu angewählt werden. Die nötigen Umcodierungen können mit Hilfe des DATA BASE MANAGEMENT SYSTEM vorgenommen werden.

In Zeile 12080 erfolgt der Sprung ins Unterprogramm zur Berechnung des Signifikanzniveaus nach der χ^2 -Ver-

-
- 1) In den Rechenzentren der Universitäten werden oft entsprechende Statistiken geführt.
 - 2) SPSS bricht an dieser Stelle den Lauf ab, da "Rückfragen" beim Benutzer nicht möglich sind.

teilung (Zeile 290), danach beginnt die Ergebnisanzeige: Nach der Überschrift mit dem Namen der geprüften Variablen erscheint eine Tabelle mit den beobachteten und erwarteten Häufigkeiten aller Kategorien, weiterhin werden angezeigt die Anzahl der Freiheitsgrade $H3-1$, der Testwert CH, auf 3 Stellen hinter dem Komma gerundet, und das Signifikanzniveau S, ebenfalls auf 3 Stellen gerundet. Den Abschluß bilden wiederum die Anzahl der gültigen Fälle und der fehlenden Fälle.

4.2.3.3 Kolmogoroff/Smirnow-1 Stichproben-Anpassungstest

Nach den üblichen Anweisungen zu Beginn startet dieser Programmteil mit der Anfrage, welche Verteilung als theoretische angenommen wird, anschließend erscheint ein kleines Menu mit den Wahlmöglichkeiten (Zeile 13010).

WELCHE VERTEILUNG WIRD ALS

THEORETISCHE VERTEILUNG ANGENOMMEN?

GLEICHVERTEILUNG.....1

NORMALVERTEILUNG.....2

POISSONVERTEILUNG....3

AUSWAHL.....

Um die Abstimmung der Menuaufteilung mit der Fehlerroutine in Zeile 61000 zu ermöglichen, wurde ein Bildschirmfenster aufgebaut mit POKE 32,4 und POKE 33,35. Bei Eingabe von falschen Zahlen außerhalb des Bereichs 1 bis 3 tritt die Fehlerabfangeinrichtung der Zeile 13020 in Aktion. Bei Eingabe von Buchstaben oder Symbolen statt Zahlen erfolgt in der Regel die Meldung REENTER durch das Betriebssystem; dies wurde jedoch durch die Zeile 61000 unterdrückt, die stattdessen eine Meldung "Falsche Eingabe. Wiederholung" kurz anzeigt, danach die falsche Eingabe löscht und zurückkehrt zum INPUT-Befehl, der zur falschen Eingabe führte. Vorherige Anweisungen über die Positionierung des Ausdrucks auf dem Bildschirm werden jedoch nicht noch einmal bearbeitet. Durch das Bildschirmfenster wird die korrekte Aufteilung des Bildschirms auch nach dieser Fehlermeldung gewährleistet.

In Zeile 13030 wird der Monitor zunächst wieder auf das normale Format erweitert für nachfolgende Anzeigen. Mit Hilfe des Unterprogramms 100 wird das Datenarray der Größe nach sortiert, dafür muß N an die subroutine übergeben werden. In der folgenden Schleife werden die Summen der X-Werte (H9) und der quadrierten X-Werte (H8) bestimmt und die Anzahl der verschiedenen realisierten Ausprägungen durch N gezählt.

Wenn eine neue Kategorie auftritt, wird N um 1 erhöht, der Wert der empirischen Verteilungsfunk-

tion der vorigen Kategorie (Y) berechnet und dem Y-Array an der Stelle N zugewiesen. Damit sind auch Verbundwerte berücksichtigt worden. Falls eine Kategorie mehrmals auftaucht (im sortierten Array geschieht dies hintereinander, wenn überhaupt), wird zu NEXT in Zeile 13040 übergegangen. Hier wird nun nach Abschluß der Schleife das arithmetische Mittel berechnet, das möglicherweise für die Anpassung an die Normalverteilung und Poissonverteilung benötigt wird. Nun muß noch der Wert der Verteilungsfunktion für die letzte Kategorie auf 1 gesetzt werden, da dies in der Schleife noch nicht geschehen ist. Erst jetzt kann der Sprung in die Unterprogramme zur Berechnung der Werte der theoretischen Verteilungsfunktion erfolgen über das ON ... GOSUB-Statement. Ist dies geschehen, erfolgt im Unterprogramm ab Zeile 860, das schon beschrieben wurde, die Bestimmung der maximalen positiven, negativen und absoluten Differenzen, der Prüfgröße Z und des Signifikanzniveaus. Es ist zu beachten, daß die Schleife des Unterprogramms mit dem Laufindex J für die Fälle nur bis N läuft und nicht bis C%. Dazu müssen Y(0) und V übergeben werden, wobei V hier ausnahmsweise nicht für die Zahl der Freiheitsgrade steht. Vielmehr muß die maximale absolute Differenz mit einem Faktor $\sqrt{V}$ multipliziert werden, um Z zu erhalten. Für den Einstichprobentest ist V gleich der Anzahl der gültigen Fälle. Die nachfolgenden Unterprogramme gestalten die Bildschirmanzeige. Das schon besprochene Unterprogramm ab Zeile 960 liefert die Anzeige des Testergebnisses.

Die Zeilen 13100 bis 13130 werden bearbeitet, falls im Menu die Option 1 gewählt wurde, also Gleichverteilung angenommen wurde. Bei Verwendung des Kolmogoroff/Smirnow-Tests sollte strenggenommen eine Schätzung von Parametern der Verteilung aus der Stichprobe nicht vor kommen. Trotzdem werden in REP-TEST zwei Möglichkeiten angeboten - entsprechend den Leistungen von SPSS.

Der Benutzer kann den Bereich spezifizieren, auf dem die untersuchte Variable gleichverteilt sein soll. Will er dies nicht, werden Minimum und Maximum der Stichprobenrealisationen bestimmt und für die Berechnungen benutzt. In Zeile 13100 erfolgt die Anfrage, ob Benutzereingabe gewünscht wird. Wenn ja, kann dies durch die Zeile 13110 geschehen.

In beiden Fällen werden in Zeile 13120 die Werte der theoretischen Verteilungsfunktion bei Gleichverteilung mit Hilfe von Minimum (H1) und Maximum (H2) berechnet, deren Werte vom Programm bestimmt wurden oder eingegeben wurden.

Zugewiesen werden die berechneten Größen dem Array X - dessen ursprüngliche Werte nun nicht mehr benötigt werden -, wiederum aber nur für die unterschiedlichen Ausprägungen, so daß das X-Array möglicherweise verkleinert und natürlich mit veränderten Werten aus diesem Unterprogramm hervorgeht.

Die Zeilen 13200 bis 13240 enthalten das entsprechende Unterprogramm bei Normalverteilungsannahme, also bei Wahl von Option 2. Es können sowohl Standardabweichung als auch Mittelwert vom Benutzer angegeben werden (Zeilen 13200 bis

13210) oder auch nur eins von beiden. Der oder die nicht spezifizierten Parameter werden aus der Stichprobe geschätzt. Auf die Problematik wurde schon hingewiesen. Ist die Standardabweichung gleich Null, erfolgt eine entsprechende Fehlermeldung.

Die Berechnung der Werte des "neuen" X-Arrays geschieht mit Hilfe der in Zeile 10 definier-ten Funktion STN(Z).

Die Berechnung der Poisson-Verteilung ist wesentlich aufwendiger als die der beiden vorhergehenden. Sie konnte aber trotzdem in den 6 Zeilen 13300 bis 13350 untergebracht werden. In der ersten Zeile erfolgt zunächst die Überprüfung, ob die Daten alle nicht - negative ganze Zahlen sind. Wenn nicht, wird das Verfahren abgebrochen und zum Menu mit der Auswahl der Testverfahren zurückgekehrt. Nun muß REP-TEST MENU gewählt werden, danach kann das DBMS benutzt werden, um offensichtlich vorhandene Fehler in den Daten zu korrigieren. Dies gilt natürlich nur, sofern nicht ein Anwenderfehler bei der Auswahl des Verfahrens vorliegt.

Sind die Daten für den Test auf Anpassung an die Poissonverteilung geeignet, kann als Folge von Zeile 13310 der Mittelwert (H9) eingegeben werden. Wird dies nicht gewünscht, erfolgt die Übernahme des in Zeile 13040 berechneten Wertes für H9. Ist der Mittelwert gleich Null, erfolgt eine entsprechende Meldung und der Abbruch des Verfahrens. Ansonsten wird in den Zeilen bis 13350 der Wert

der Poissonverteilung für die verschiedenen Stichprobenrealisationen der Untersuchungsvariablen berechnet. Um den Rechenaufwand so gering wie möglich zu halten, wurden folgende Überlegungen ausgenutzt: Falls X poissonverteilt ist, so ist

$$P(X \leq x) = \sum_{k=0}^x P(X=k) = \sum_{k=0}^x e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \cdot \sum_{k=0}^x \frac{\lambda^{k-1}}{(k-1)!} \cdot \frac{\lambda}{k} \quad \text{für } x > 0 ,$$

d. h. jeder Summand (H3) geht aus dem vorhergehenden durch Multiplikation mit λ/k hervor. Startwert ist $H3 = \frac{\lambda^0}{0!} = 1$. Erst die Summe (H4) wird anschließend mit $e^{-\lambda}$ (H1) multipliziert. Weiter ist folgendes zu bedenken: Hat man 2 Stichprobenwerte x_1 und x_2 mit $x_1 < x_2$, so ist:

$$P(X < x_2) = \sum_{k=0}^{x_2} e^{-\lambda} \cdot \frac{\lambda^k}{k!} = \sum_{k=0}^{x_1} e^{-\lambda} \cdot \frac{\lambda^k}{k!} + \sum_{k=x_1+1}^{x_2} e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \cdot \left(\sum_{k=0}^{x_1} \frac{\lambda^k}{k!} + \sum_{k=x_1+1}^{x_2} \frac{\lambda^k}{k!} \right) .$$

Somit kann für jeden Wert (ab dem zweiten) die Summe H4 des vorhergehenden benutzt werden. Aus diesem Grund wird H4 innerhalb der J-Schleife in Zeile 1334Ø nicht auf Null gesetzt, sondern weiter in die Rechnung einbezogen. So konnte der Rechenaufwand gering gehalten werden, da überflüssige Potenzierungen und Summationen entfallen. Nachdem das veränderte Array X aufgebaut ist, erfolgt die Rückkehr in Zeile 13Ø4Ø.

4.2.3.4 χ^2 -Homogenitätstest

Bei diesem Verfahren erfolgt vor der Eingabe der Variablennamen ein Hinweis auf die Codierung, die das Programm benötigt. Falls diese nicht eingehalten ist, wird in Zeile 14010 nach einer Fehlermeldung abgebrochen. Das Verfahren benötigt insgesamt 9 Zeilen (14000 bis 14080), da mehrere Fälle unterschieden werden müssen. Zunächst werden in 14020 bis 14030 die gemeinsamen Häufigkeiten der beiden Variablen, die Randhäufigkeiten und die Maxima (H1% und H2%) der Merkmalswerte bestimmt. Die Anzahl der Freiheitsgrade V ergibt sich als $H1\% \cdot H2\%$, da von Null aufwärts codiert werden sollte. Wenn $V=1$ ist, so liegt der Fall einer Vierfeldertabelle vor. Dann wird für eine Anzahl von gültigen Fällen bis einschließlich 30 ein Fisher-Test über das Unterprogramm 770 gerechnet (Zeile 14040). Für mehr als 30 gültige Fälle wird für $V=1$ ein χ^2 -Test durchgeführt, die Berechnung der Prüfgröße CH muß dann aber die Yates'sche Korrektur enthalten, wie sie in Zeile 14050 zu finden ist.

Zeile 14060 enthält die Berechnung des CH-Wertes für $V>1$.

In Zeile 14070 stehen die Anweisungen für die Berechnung des Signifikanzniveaus (GOSUB 260) mit Hilfe der χ^2 -Verteilung und die Bildschirmanzeige für das Testergebnis, falls kein Fisher-Test durchgeführt wurde.

Die letzte Zeile (14080) wird in allen 3 Fällen abgearbeitet.

4.2.3.5 Fisher-Test

Das Verfahren benötigt nur 3 Zeilen, da die Berechnungen und die Bildschirmanzeige zum größten Teil über Unterprogramme erfolgen. In Zeile 15000 findet sich die übliche Befehlsfolge zu Anfang eines Testverfahrens und der Hinweis darauf, daß die Variablen 0/1 codiert sein müssen. Diese Codierung wird nun in Zeile 15010 überprüft. Das Unterprogramm 720 prüft die Werte im Y-Array, deshalb muß für die Kontrolle des X-Arrays dieses in das Y-Array kopiert werden, das in das A-Array verschoben wird. A(0,J) enthält keine Datenwerte. Dies ist wichtig, da andere Prozeduren später ja möglicherweise auf dasselbe Eingabearray A zurückgreifen. Gerechnet wird aber nur mit A(I,J), I>0.

Falls die Codierung in Ordnung ist, aber die Zahl der gültigen Fälle 30 übersteigt, wird ein χ^2 -Homogenitätstest durchgeführt (Sprung zur Zeile 14010) nach einer entsprechenden Meldung. Das Unterprogramm 830 dient dazu, die Häufigkeiten - bivariate und univariate - auszu-zählen, anschließend wird auf dem Bildschirm eine Tabelle angezeigt (GOSUB 840). Das Unterprogramm 770 berechnet Fisher's exakte Wahrscheinlichkeit, das einseitige Signifikanzniveau. Mit der üblichen Anweisungsfolge schließt diese letzte Zeile 15020.

4.2.3.6 Kolmogoroff/Smirnow-2 Stichproben-Test

Dieser Test beansprucht die Zeilen 16000 bis 16040; er ist wesentlich kürzer als der Einstichprobentest, da keine Eingaben vom Benutzer erwartet werden und die Unterprogramme zur Bestimmung der theoretischen Verteilungsfunktion entfallen. Allerdings ist die Bestimmung der empirischen Verteilungsfunktionen etwas aufwendiger, insbesondere wenn wegen des doch stark begrenzten Speicherplatzes nicht zuviele Arrays dimensioniert werden können. Die Datenarrays X und Y müssen wiederum für die beiden zu vergleichenden Verteilungsfunktionen benutzt werden. Y dient zunächst als Gruppierungsvariable zur Festlegung der beiden Stichproben, seine Werte sollen deshalb alle nur 0 oder 1 sein. Nach der üblichen Initialisierung in Zeile 16000 wird zunächst das X-Array der Größe nach sortiert, in der folgenden Schleife werden die Stichprobenumfänge der beiden Stichproben durch T%(0) und T%(1) gezählt. Da diese beiden Variablen später verändert werden, ihre Werte für den Ausdruck aber noch gebraucht werden, müssen sie an H8% und H9% übergeben werden. Es wird nun der Code der größeren der beiden Gruppen (H1%) bestimmt und ihr Umfang (N). Q ist also der Umfang, H2% der Code der kleineren Stichprobe.

In einer Schleife mit negativer Schrittweite werden nun in den Zeilen 16020 bis Anfang 16040 die empirischen Verteilungen der beiden Stichproben bestimmt, ihre Werte den Arrays X und Y

zugeordnet. Wobei Y die kumulierte Verteilungsfunktion der größeren Stichprobe enthält. Hierbei wird nicht mit reduzierten Arrays gearbeitet. Da für Werte der ersten Stichprobe auch die Verteilungsfunktion der zweiten berechnet werden muß (und umgekehrt), die eigentlichen Daten aber nach der Neuzuweisung nicht mehr zugänglich sind, können die Stichprobenwerte nur noch indirekt anhand des Index J , der ihre Stellung in der geordneten Reihe der Datenpunkte angibt, identifiziert werden. Deshalb müssen die Arrays ihren ursprünglichen Umfang behalten (bzw. ihr relevanter Bereich muß gleich bleiben), damit sichergestellt ist, daß beim Vergleich der Verteilungsfunktionen $Y(J)$ und $X(J)$ die Funktionswerte desselben Arguments verglichen werden.

Wenn zwei gleiche Werte im ursprünglichen X -Array aufeinanderfolgen, so werden für den zweiten die Funktionswerte des vorigen übernommen und der Zähler der Gruppe, aus der der aktuelle Wert stammt, um 1 reduziert.

Tritt ein neuer Wert auf, so werden die Funktionswerte mit Hilfe des sich ständig verändernden Zählers ihrer Gruppe ($T\%(H1\%)$ bzw. $T\%(H2\%)$) berechnet und der Zähler der Gruppe, aus der der Wert stammt, wird um 1 reduziert. So ist sichergestellt, daß die Funktionen bei mehrfach aufgetretenen Argumenten denselben Wert haben und daß Argumente, die in einer Gruppe nicht realisiert wurden, den Funktionswert des nächst-kleineren Elementes der Gruppe zugewiesen bekommen.

Nach Abschluß der Schleife werden die Parameter $A\%$, N und V an das nachfolgend aufgerufene Unter-

programm 86Ø übergeben. Die Belegung von A%=3 ist notwendig, da im Einstichprobenfall für jeden Stichprobenwert 2 Differenzen berechnet werden, nämlich $Y(J)-X(J)$ und $Y(J-1)-X(J)$, außer für die Option 3, die Poissonverteilung. Für den Zweistichprobenfall soll ebenfalls nur die Differenz $Y(J)-X(J)$ berechnet werden, die Bildung der zweiten kann aber bei A%=3 übersprungen werden, wie aus der entsprechenden subroutine zu ersehen ist. Die Bildschirmanzeige entspricht im wesentlichen der des Einstichprobenfalles, ist aber erweitert worden um die Zeile, die die Stichprobenumfänge der beiden Gruppen anzeigt.

4.2.3.7 Wald/Wolfowitz-Runtest

Für dieses Verfahren in den Zeilen 17ØØØ bis 1713Ø werden zwei Stichproben benötigt. Die Variable 2 dient als Gruppierungsvariable und sollte deshalb 0/1 codiert sein. Eine entsprechende Anzeige wird durch Zeile 17ØØØ ausgelöst. Dort sind außerdem die üblichen Anweisungen zur Vorbereitung der Rechnung zu finden. Die Anzahl der Runs ist gleich der Anzahl der Signwechsel bzw. Gruppenwechsel in der geordneten Reihe der gepoolten Stichprobe aus $X(J)$ plus 1. Deshalb wird in Zeile 17Ø1Ø nach den üblichen Unterprogrammen (72Ø, 67Ø, 61Ø) zunächst einmal der X/Y-Sorter (1ØØ) eingesetzt, danach werden in einer Schleife die 1-Codierungen gezählt (durch H2%) und es wird geprüft, ob Vorzeichenwechsel auftreten. Ist dies der Fall, wird die Zählvariable H9% um eins erhöht. Anschließend

wird geprüft, ob dieser Wechsel in einem Verbundwertebereich des X-Arrays aufgetreten ist. Bei Verbundwerten sollte der Wald/Wolfowitz-Test gar nicht benutzt werden, zumindest nicht, wenn die Verbundwerte aus verschiedenen Gruppen stammen. Das Flag H4% wird beim ersten Auftreten eines solchen unerwünschten Falles auf 1 gesetzt, damit später eine entsprechende Bildschirmanzeige erfolgen kann.

In 17020 wird H9% nochmals um 1 erhöht und enthält nun die Anzahl der Runs, H1% wird mit der Anzahl der O-Codierungen belegt. Nach dem Ausdrucken der Überschrift werden zunächst die beiden Stichprobenumfänge angezeigt, danach die Anzahl der Runs. Falls die Summe von H1% und H2% (diese entspricht der Zahl der gültigen Fälle C%) kleiner oder gleich 30 ist, erfolgt die Bestimmung des Signifikanzniveaus nach der exakten Verteilung der Prüfgröße H9%. In diesem Fall wird anstelle des Testwerts Z nur ein Strich gedruckt als Zeichen dafür, daß S exakt berechnet wurde. Dies erfolgt dann in den Zeilen 17100 bis 17130.

Falls $C > 30$ ist, wird die Verteilung von H9% durch die Normalverteilung approximiert, deren Parameter in den Zeilen 17030 und 17040 berechnet werden. Bei einer Standardabweichung von Null erfolgt nach einer Fehlermeldung der Abbruch des Verfahrens und die Rückkehr zum Menu. Nun wird Z berechnet, für $C < 50$ mit einer Stetigkeitskorrektur des Zählers von 0,5. Zeile 17050 enthält die Bestimmung von S und den Ausdruck des Wertes Z.

In Zeile 17060 folgen die Anweisungen für den restlichen Teil der Bildschirmanzeige, der für beide Berechnungsmethoden des (einseitigen) Signifikanzniveaus S gleich ist. Falls das Flag $H4$ nicht gleich Null ist, also Verbundwerte zwischen den beiden Gruppen aufgetreten sind, erfolgt eine entsprechende Warnung, da das Testergebnis mit Vorsicht interpretiert werden muß.

Die Zeilen 17100 bis 17130 ermöglichen die Berechnung des Signifikanzniveaus über die exakte Verteilung von $H9$, die auf einigen Umformungen der in der der Literatur benutzten Formeln für diese Verteilung beruht; diese lauten:

$$P(X \leq x) = \sum_{k=2}^x P(X = k) \quad \text{mit}$$

$$P(X = k) = \binom{n_1+n_2}{n_1}^{-1} \cdot 2 \cdot \binom{n_1-1}{k/2-1} \cdot \binom{n_2-1}{k/2-1}, \text{ wenn } k \text{ gerade ist,}$$

und

$$P(X = k) = \binom{n_1+n_2}{n_1}^{-1} \cdot \left[\binom{n_1-1}{\frac{k+1}{2}-1} \cdot \binom{n_2-1}{\frac{k+1}{2}-2} + \binom{n_1-1}{\frac{k+1}{2}-2} \cdot \binom{n_2-1}{\frac{k+1}{2}-1} \right],$$

wenn k ungerade ist.

Man sieht, daß in beiden Formeln k nur in der Form $k/2$ bzw. $(k+1)/2$ auftaucht; dies wird nun zur Vereinfachung der Formeln ausgenutzt. Die Summe für $P(X \leq x)$ wird aufgeteilt in eine Summe mit geraden und eine mit ungeraden Indizes. Es ist:

$$\sum_{k=2}^x P(X = k) = \sum_{k \text{ ger.}} \binom{n_1+n_2}{n_1}^{-1} \cdot 2 \cdot \binom{n_1-1}{k/2-1} \cdot \binom{n_2-1}{k/2-1}$$

$$= \sum_{j=1}^{x'} \binom{n_1+n_2}{n_1}^{-1} \cdot 2 \cdot \binom{n_1-1}{j-1} \cdot \binom{n_2-1}{j-1} ,$$

wobei $x' = x/2$, falls x gerade ist, und

$x' = (x-1)/2$, falls x ungerade ist.

Weiter gilt:

$$\begin{aligned} \sum_{\substack{k=3 \\ k \text{ ung.}}}^x P(X = k) &= \sum_{\substack{k=3 \\ k \text{ ung.}}}^x \binom{n_1+n_2}{n_1}^{-1} \cdot \left[\binom{n_1-1}{\frac{k+1}{2}-1} \cdot \binom{n_2-1}{\frac{k+1}{2}-2} \right. \\ &\quad \left. + \binom{n_1-1}{\frac{k+1}{2}-2} \cdot \binom{n_2-1}{\frac{k+1}{2}-1} \right] \\ &= \sum_{\substack{k=3 \\ k \text{ ung.}}}^x \binom{n_1+n_2}{n_1}^{-1} \cdot \left[\binom{n_1-1}{\frac{k-1}{2}} \cdot \binom{n_2-1}{\frac{k-3}{2}} + \binom{n_1-1}{\frac{k-3}{2}} \cdot \binom{n_2-1}{\frac{k-1}{2}} \right] \\ &= \sum_{j=1}^{x''} \binom{n_1+n_2}{n_1}^{-1} \cdot \left[\binom{n_1-1}{j} \cdot \binom{n_2-1}{j-1} + \binom{n_1-1}{j-1} \cdot \binom{n_2-1}{j} \right] , \end{aligned}$$

wobei $x'' = (x-1)/2$, falls x ungerade ist, und

$x'' = (x-2)/2 = x/2 - 1$, falls x gerade ist.

Für ungerades x ist $x'=x''$, für gerades x ist $x''=x'-1$. Dies ist selbstverständlich, da für gerades x die erste Summe einen Summanden mehr enthalten muß. Setzt man nun $\hat{x}=x''$ für beide Fälle, d. h. $\hat{x}=\text{INT}((x-1)/2)$, so kann man folgendermaßen zusammenfassen:

$$P(X \leq x) = \binom{n_1+n_2}{n_1}^{-1} \cdot \sum_{j=1}^{\hat{x}} \left[2 \cdot \binom{n_1-1}{j-1} \cdot \binom{n_2-1}{j-1} + \binom{n_1-1}{j} \cdot \binom{n_2-1}{j-1} + \binom{n_1-1}{j-1} \cdot \binom{n_2-1}{j} \right] + z \cdot 2 \cdot \binom{n_1-1}{x/2-1} \cdot \binom{n_2-1}{x/2-1}$$

wobei $z = 0$, falls x ungerade ist, und $z = 1$ gilt, falls x gerade ist.

Im Programm entspricht dem Wert $\hat{x}$ die Variable H8%, n_1-1 und n_2-1 werden durch H6% und H7% dargestellt, z durch H3%. Die mehrfach auftauchenden Summanden werden nur einmal berechnet, d. h. z. B.: $\binom{n_1-1}{j-1}$ aus dem ersten Summanden wird in H6 zwischengespeichert, damit es für den dritten Summanden nicht neu berechnet werden muß. So braucht die subroutine 31Ø zur Berechnung der Binomialkoeffizienten nur viermal angesprungen zu werden. Nach Ende der Berechnung wird zu Zeile 17Ø6Ø gesprungen für die Bildschirmanzeige.

4.2.3.8 McNemar-Test

Das in den Zeilen 18000 bis 18500 gerechnete Verfahren benötigt zwei verbundene Stichproben, deren beider Werte 0/1 codiert sein müssen. Zur Überprüfung muß das Y-Array nach $A(\emptyset, J)$ verschoben werden und das X-Array nach $Y(J)$. In Zeile 18010 werden nun die Häufigkeiten ausgezählt und die Variable H3% mit der Summe der Elemente der Nebendiagonalen der 2x2-Tabelle belegt. Das zweiseitige Signifikanzniveau S wird mit Hilfe der Binomialverteilung berechnet, falls die Anzahl der "Wechsler" kleiner oder gleich 10 ist (Zeilen 18030 und 18040). Ist $H3\% > 10$, so wird in Zeile 18020 der Testwert CH mit der Yateskorrektur berechnet, über die χ^2 -Verteilung S bestimmt (Unterprogramm 260) und zum Beginn der Bildschirmanzeige in 18050 gesprungen. Das Programm liefert eine Tabelle mit den Zellenbesetzungen der Vierfeldertabelle und für $H3\% > 10$ den Testwert CH. Anschließend wird das Signifikanzniveau angezeigt und die Anzeige in üblicher Weise abgeschlossen. Die Zeile 18500 stellt eine kleine subroutine dar, die zur Gestaltung der Tabelle benötigt wird.

4.2.3.9 Cochran-Q-Test

Dieses Verfahren beansprucht 6 Programmzeilen. Wegen der Bildschirmanzeige wurde die Zahl der verbundenen Stichproben auf maximal fünf beschränkt. Die Untersuchungsvariable muß in allen

Stichproben 0/1-codiert sein. In Zeile 19030 wird die Anzahl der 1-Codierungen für jede Stichprobe (T%) und für jeden Fall (X) aus gezählt; das X-Array wird hier nicht für Daten gebraucht und kann deshalb für das Auszählen verwendet werden. Weiter werden in dieser Zeile die für die Prüfgröße benötigten Summen der X(J) und der quadrierten X(J), der T%(I) und der quadrierten T%(I) berechnet. Falls der Nenner der Prüfgröße (H1) nicht positiv ist, werden S und CH auf den Wert 999(IC%) gesetzt. Andernfalls kann CH in Zeile 19040 berechnet und S über das Unterprogramm 260 bestimmt werden.

Zunächst werden für die Bildschirmanzeige die Namen der Testvariablen (die den Meßwertreihen entsprechen) untereinander aufgelistet, danach wird dieser Teil des Monitors "gesperrt" durch POKE-Befehle und rechts neben die Variablen namen die Häufigkeiten für die beiden Codes der Untersuchungsvariablen geschrieben. Danach wird der Bildschirm wieder zur vollen Breite erweitert und es werden die Anzahl der Freiheitsgrade, die Größe CH und die Signifikanz angezeigt. Schon bei der Besprechung der Unterprogramme zur Variablenidentifikation und für den Ladeprozeß wurde auf eine Besonderheit dieses Verfahrens hingewiesen. Die subroutine zum Aussortieren der missing values (Zeile 540) verändert nämlich das Eingabearray A(I,J). Aussortiert werden alle Fälle, bei denen mindestens eine Variable auf missing value steht. Bei einer anderen Variablenzusammenstellung können andere Fälle als fehlend erkannt werden als bei der gegebenen. Deshalb dürfen nach einem

Cochran-Test die Eingabearrays A(I,J) nicht mehr für weitere Testverfahren verwendet werden. Das bedeutet, daß der CPU-Katalog gelöscht werden muß. Dies geschieht durch die letzten Statements der Zeile 19060 nach Abschluß der Ergebnisanzeige durch das Unterprogramm 650. Bei einem unmittelbar anschließend durchgeführten Test arbeitet der Rechner so, als ob noch keine Variablen geladen worden seien.

4.2.3.10 Runtest

Der Runtest wird nur in seiner abgeleiteten Form zur Untersuchung einer Stichprobe auf Zufälligkeit der Auswahl durchgeführt. Dazu kann wahlweise am Median oder am arithmetischen Mittel dichotomisiert werden. Die Auswahl wird in Zeile 20010 getroffen durch die Antwort auf die Frage "Soll am Median dichotomisiert werden?".

Wird mit "N" geantwortet, wird in Zeile 20020 das arithmetische Mittel berechnet. Ansonsten wird in Zeile 20030 der Median mit Hilfe des Unterprogramms 800 bestimmt.

H7 ist jedenfalls der sogenannte "Cutpoint"-entweder Median oder arithmetisches Mittel. In Zeile 20040 finden sich die Anweisungen zur Bestimmung der Anzahl von Runs, die sich bis zur Zeile 20050 erstrecken.

Es wird durch T% die Anzahl der negativen und der nicht-negativen Abweichungen vom Cutpoint gezählt. H1% enthält jeweils das Vorzeichen für das vorher behandelte Element des X-Arrays, H2%

das des aktuellen. Sind beide Variablen verschieden, liegt ein Vorzeichenwechsel vor und der Zähler H3% wird um 1 erhöht. Zu Beginn der Zeile 20050 wird die Variable für das Vorzeichen der vorhergehenden Abweichung aktualisiert. Nach dem Durchlaufen der Schleife gibt H3% die Anzahl der Runs an, da der Startwert 1 in der Schleife um die Anzahl der Vorzeichenwechsel erhöht wurde. Die Verteilung von H3% wird über die Normalverteilung approximiert,¹⁾ für C%≤50 wird eine Stetigkeitskorrektur berücksichtigt. H3 und H4 sind die Parameter der Normalverteilung (Zeile 20050 und 20060). Zeile 20070 enthält die gesamte Bildschirmanzeige, die die Anzahl der Runs, die Anzahl der negativen und nicht-negativen Abweichungen vom Cutpoint, den Wert Z und das zweiseitige Signifikanzniveau S umfaßt.

4.2.3.11 χ^2 -Unabhängigkeitstest

Für die beiden Variablen dieses Verfahrens, das die Zeilen 21000 bis 21080 einnimmt, wird vorausgesetzt, daß sie ganzzahlig, von 0 bis höchstens 5, codiert sind. Ein entsprechender Hinweis wird vor der Variableneingabe angezeigt (Zeile 21000).

In den Zeilen 21010 bis 21025 werden die Maxima der beiden Variablen in der Stichprobe bestimmt und es wird die Codierung überprüft. Zeile 21030 leistet die Auszählung der bivariaten Häufigkeiten und die Berechnung der Randhäu-

1) Dies entspricht dem Vorgehen von SPSS. Vgl. dazu Norusis (1979), S. 117

figkeiten. In Zeile 21035 werden die erwarteten Häufigkeiten (bei Unabhängigkeit) berechnet und die Prüfgröße CH.

Durch die Anweisungen der Zeile 21040 werden die Anzahl der Freiheitsgrade, die Prüfgröße CH und die Signifikanz S angezeigt. Anschließend kann auf Wunsch eine Häufigkeitstabelle angezeigt werden (durch die Zeilen 21070 und 21080), wonach der Ausdruck des Testergebnisses durch Zeile 21040 noch einmal wiederholt wird.

Wird in Zeile 21050 mit "N" geantwortet, wird das Verfahren auf die bekannte Weise abgeschlossen.

4.2.3.12 Spearman

Der Test benötigt nur sechs Zeilen. Er arbeitet mit zwei Variablen, deren Unabhängigkeit geprüft werden soll; ihre Codierung ist beliebig. Beide Variablen werden zunächst getrennt sortiert und mit Rangplätzen versehen. Da die subroutines für diese Aufgaben das X-Array bearbeiten, müssen in Zeile 22010 die beiden Datenreihen vertauscht werden, bevor für die zweite Variable die Rangplätze zugewiesen werden können.

In der Schleife in Zeile 22020 werden die Rangplatzdifferenzen quadriert und aufsummiert. Für die Berechnung der Prüfgröße braucht man außerdem noch den Spearman'schen Korrekturfaktor für beide Variablen, der durch die Rankingroutine schon jeweils mit berechnet wurde und als H9 bzw. H3 zur Verfügung steht. Falls einer dieser beiden Werte mit $N^3 - N$ übereinstimmt (wobei $N=C\%$, d. h.

gleich der Anzahl der gültigen Fälle ist), kann die Prüfgröße nicht berechnet werden und S und Z werden auf 999 gesetzt.

Andernfalls wird die Prüfgröße Z in Zeile 22030 berechnet und - über die t-Verteilung mit N-2 Freiheitsgraden in Unterprogramm 210 - das zweiseitige Signifikanzniveau S.

Die Anzeige, die durch die Zeilen 22040 und 22050 hervorgerufen wird, enthält die Anzahl der Ties für beide Variablen, den Korrelationskoeffizienten, die Anzahl der Freiheitsgrade, den Wert der Prüfgröße T und das Signifikanzniveau S.

4.2.3.13 Bravais/Pearson

Der Test beansprucht nur vier Zeilen, 23000 bis 23030; in Zeile 23010 werden zunächst die für den Korrelationskoeffizienten benötigten Summen bestimmt, dann der Nenner H6 berechnet. Ist H6 gleich Null, kann der Korrelationskoeffizient nicht berechnet werden. Nach einer Fehlermeldung wird das Verfahren abgebrochen.

Ansonsten werden in Zeile 23020 Z bestimmt, die Prüfgröße T und die Anzahl der Freiheitsgrade. T und V sind Eingänge des Unterprogramms 210. Die Ergebnisanzeige besteht aus dem Korrelationskoeffizienten, der Anzahl der Freiheitsgrade, dem Wert für T und dem zweiseitigen Signifikanzniveau S.

4.2.3.14 Mediantest

Dieser Test beansprucht wesentlich mehr Platz als der vorhergehende, nämlich 17 Zeilen, von Zeile 24000 bis 24220. Die Gruppierung der Stichproben erfolgt durch Variable 2. Falls diese Variable Y nicht ganzzahlig codiert ist, wird nach einer Fehlermeldung abgebrochen (Zeile 24010). In den Zeilen 24020 bis 24030 werden Minimum (H1%) und Maximum (H2%) von Y bestimmt. Falls die Anzahl der Stichproben größer ist als 6, wird nicht weitergearbeitet (Zeile 24040).

In Zeile 24050 beginnen die Rechnungen. Es wird zunächst der Median bestimmt (800), dabei wird die Variable X umcodiert. Sie wird 0, falls $X(J)$ kleiner oder gleich dem Median ist, und 1, falls $X(J)$ größer ist als der Median. Danach werden die Häufigkeiten dieser neuen X-Werte in den verschiedenen Stichproben durch N% gezählt und die Randhäufigkeiten bestimmt. S gibt den Umfang der Gruppen an und T% die Anzahl aller Werte kleiner gleich bzw. größer als der Median. Die Anweisungen GOSUB 850 und GOSUB 840 lassen eine Tabelle auf dem Bildschirm erscheinen mit den genannten Häufigkeiten. Es müssen nun 3 Fälle unterschieden werden:

a) $H3\% = 1$ und $C \leq 30$

Es handelt sich um eine Vierfeldertabelle, deren Gesamtumfang höchstens 30 Fälle beträgt. Über das Unterprogramm 770 wird ein Fisher-Test durchgeführt.

b) $H3\% = 1$ und $C\% > 30$

Es handelt sich um eine Vierfeldertabelle, doch ist die Anzahl der Fälle ausreichend groß für einen χ^2 -Test. In den Zeilen 24100 und 24110 wird die nach Yates korrigierte Prüfgröße CH berechnet und das Signifikanzniveau S (260). Es schließt sich die Ergebnisanzeige an.

c) $H3\% > 1$

Es liegen mehr als zwei Ausprägungen der Variablen Y vor, d. h. es gibt mehr als zwei Stichproben.

In Zeile 24070 wird die Prüfgröße CH ohne Korrektur berechnet. Die Anzahl der Zellen mit erwarteten Häufigkeiten kleiner als fünf wird durch H8% gezählt und falls eine erwartete Häufigkeit kleiner als eins ist, wird ein Flag Q auf 1 gesetzt. Zeile 24090 enthält wiederum den Sprung ins Unterprogramm 260 zur Berechnung des Signifikanzniveaus. Bevor jedoch die Bildschirmanzeige erweitert wird, erfolgt noch die Überprüfung der Indikatoren für nicht eingehaltene Voraussetzungen des Tests, H5 und Q. Sind H5 und Q beide Null, wird sofort zur Anzeige in Zeile 24120 gesprungen.

Ist einer von beiden positiv, wird die subroutine in den Zeilen 24200 bis 24220 abgearbeitet, die Hinweise anzeigt, falls mindestens eine erwartete Häufigkeit kleiner als eins ist und/oder mehr als 20% der erwarteten Häufigkeiten kleiner als fünf sind. Auf Tastendruck werden diese Hinweise gelöscht und es erfolgen die Anzeigen aus den Zeilen 24120 und 24130: die Anzahl der Freiheitsgrade, der Wert CH und das Signifikanzniveau

S. Die Anzahl der gültigen Fälle bildet wie immer die letzte Zeile der Anzeige.

4.2.3.15 Mann/Whitney-U-Test

Das Verfahren arbeitet mit zwei unabhängigen Stichproben, die durch die Gruppierungsvariable Y gebildet werden. Diese Variable 2 muß also 0/1 codiert sein. Die gepoolte Stichprobe wird sortiert und mit Rangplätzen versehen (GOSUB 100: GOSUB 180 in Zeile 25020). Dann wird in der J-Schleife die Summe der Rangplätze für beide Gruppen berechnet und ihr Umfang gezählt. Anschließend wird die Prüfgröße U, hier mit H1 bezeichnet, berechnet und gegebenenfalls durch U' ersetzt.

In Zeile 25030 werden der Testwert Z und das zweiseitige Signifikanzniveau S berechnet und der Ausdruck begonnen. Für Z wird der Korrekturfaktor nach Spearman für das Auftreten von ties berücksichtigt.

Das Unterprogramm 750 druckt für beide Gruppen den Code, die Anzahl und den mittleren Rang.

Zeile 25040 liefert außerdem die Anzeige für U, die Anzahl der ties, den Testwert Z und die Signifikanz.

4.2.3.16 Kruskal/Wallis-H-Test

Da dieses Verfahren im Ablauf dem von Mann/Whitney ähnlich ist, sind auch die entsprechenden Programmteile fast gleich aufgebaut. Zugelassen sind bis zu 6 Kategorien der Gruppierungsvariable, die von 0 aufwärts codiert sein muß. Wie im Test zuvor werden zunächst die Rangsummen aller Gruppen berechnet und ihr Umfang ausgezählt. Die Prüfgröße H , hier mit CH bezeichnet, wird in 26020 ausgerechnet und in 26030 um die Tie-Korrektur nach Spearman erweitert. Das Signifikanzniveau S wird über die χ^2 -Verteilung (Unterprogramm 260) bestimmt. Der Ausdruck ist dem des Mann/Whitney-U-Tests gleich bis auf die Anzahl der verschiedenen Codes, für die die Häufigkeiten und die mittleren Ränge angezeigt werden.

4.2.3.17 Dixon/Mood-Vorzeichen-Test

Dieser Test benötigt die Zeilen 27000 bis 27070. Zunächst werden - nach der üblichen Einleitung - in Zeile 27010 und 27020 die Vorzeichen der Differenzen der beiden Meßwertreihen bestimmt. Positive und negative Vorzeichen werden gezählt; falls die beiden Meßwerte gleich sind, ist das Paar nicht von Interesse. Die Summe dieser beiden Zahlen $H1\%$ (für + Differenzen) und $H2\%$ (für - Differenzen) ist $H3\%$, der "reduzierte Stichprobenumfang". Die Variable K wird mit dem Minimum von $H1\%$ und $H2\%$ belegt (Zeile 27030).

Falls die Anzahl der interessierenden Paare (H3%) nicht größer als 25 ist, wird das zweiseitige Signifikanzniveau über die Binomialverteilung mit $\pi = 0,5$ berechnet und ein Flag H7% auf 1 gesetzt für die Steuerung der Anzeige (Zeile 27040).

Wenn H3% 25 übersteigt, wird die Verteilung durch die Normalverteilung approximiert, deren Funktionswert Z in Zeile 27050 errechnet wird. Zu diesem Zweck wird K auf das Maximum von H1% und H2% gesetzt. Danach wird S berechnet.

Der Ausdruck enthält die Anzahl der positiven und der negativen Differenzen. Der Wert von Z wird nur für H3% 25 berechnet, also auch nur dann angezeigt. Erreicht wird dies durch die Überprüfung des Flags H7%. Wenn S über die Binomialverteilung bestimmt wurde, steht H7% auf 1, dann wird anstelle von Z nur ein Stich angezeigt. Dadurch wird klar, daß die exakte Wahrscheinlichkeit dem Signifikanzniveau zugrunde liegt.

4.2.3.18 Wilcoxon-Test für verbundene Paare

Das Vorgehen beim Wilcoxon-Test ist anders als beim einfachen Vorzeichentest - trotz gewisser Parallelen zwischen beiden Verfahren. Relevant sind ebenfalls nur die Paare mit Differenzen ungleich Null. Diese werden in Zeile 28020 durch H9% gezählt. Außer dem Vorzeichen interessiert aber auch die Größe der Abweichung.

Das Array X wird deshalb neu belegt mit den absoluten Werten der Differenzen, dem Array Y werden die zugehörigen Vorzeichen zugewiesen - jeweils nur für die Paare mit nicht verschwindenden Differenzen. Der relevante Bereich der beiden Arrays wird also möglicherweise kleiner. Außerdem wird die Häufigkeit positiver Abweichungen gezählt durch H1%. In Zeile 28030 wird nun das X-Array der absoluten Differenzen sortiert und mit Rangplätzen versehen. Dann wird für die beiden durch "+" und "-" gekennzeichneten Gruppen jeweils die Rangsumme berechnet. Wenn nur positive oder nur negative Differenzen auftreten, kann der Test nicht gerechnet werden, Z und S werden auf 999 gesetzt. H2% bezeichnet die Anzahl der negativen Abweichungen, die Variable H3 wird mit dem Minimum der beiden Rangsummen $S(0)$ und $S(1)$ belegt (Zeile 28040). In Zeile 28050 folgt die Berechnung der Prüfgröße Z und des zweiseitigen Signifikanzniveaus.

Die Ergebnisanzeige ähnelt der des einfachen Dixon/Mood-Vorzeichen-Tests, enthält jedoch zusätzlich die mittleren Ränge für die Gruppe mit den positiven und die mit den negativen Abweichungen.

2.4.3.19 Friedman-Test

Das Verfahren konnte auf sechs Programmzeilen beschränkt werden (29000 bis 29050); es verlangt eine Eingabe für die Anzahl der Gruppen (VC%).

In Zeile 29020 beginnt die Rechnung. Für jeden Fall werden die Werte der einzelnen Messungen dem X-Array zugewiesen, das Y-Array erhält eine Identifikationsnummer für die ursprüngliche Reihenfolge der Werte. Dann wird das X-Array sortiert und mit Rangplätzen versehen. Danach werden die Rangsummen der verschiedenen Gruppen (oder Meßwertreihen) um die entsprechenden Rangplätze des aktuellen Falles erhöht. In Zeile 29030 wird durch P die Summe aller quadrierten Rangsummen aller Meßwertreihen gebildet.

Nun kann die Berechnung der Prüfgröße CH erfolgen; V ist um eins kleiner als die Anzahl der Gruppen- bzw. Meßwertreihen. Nach dem Sprung ins Unterprogramm 260 zur Berechnung von S beginnt in Zeile 29040 der Ausdruck.

Neben den Namen der Testvariablen werden die mittleren Ränge der entsprechenden Meßwertreihen angezeigt unter Benutzung des Bildschirmfensters, danach erfolgt der übliche Ausdruck für die Freiheitsgrade, den Testwert CH und das Signifikanzniveau.

Den Abschluß bildet in Zeile 29050 die Korrektur des CPU-Variablenkatalogs. Bei der Erläuterung zum Programmteil für den Cochran-Q-Test wurde dieser "Löschvorgang" ausführlich begründet. Hier soll nur noch einmal in Erinnerung gerufen werden, daß nach Durchführung eines Friedman-Tests für den Rechner keine Variable mehr in der CPU vorhanden ist, da der Zähler und die Identifikationsnummern aller CPU-Variablenpositionen auf Null gesetzt werden.

4.3 Programm "TEST MIT HILFE" - Auswahl der Tests durch das Programm

Dieses Programm ist für Anwender statistischer Verfahren gedacht, die noch Schwierigkeiten haben bei der Entscheidung, welches Verfahren für ihr aktuelles Problem geeignet ist und sich deshalb in den herkömmlichen Statistikprogrammpaketen nicht zurechtfinden. Dabei sind ihnen die Entscheidungskriterien meist bekannt, aber die Zusammenführung der verschiedenen Parameter zu einer Entscheidung macht Probleme. Für diesen Anwenderkreis werden schrittweise die benötigten Parameter erfragt und dann wird vom Programm die Entscheidung über den durchzuführenden Test getroffen.

Um dies zu ermöglichen, war eine klare Strukturierung des Problemkreises erforderlich, denn die Auswahl der Verfahren muß eindeutig sein, wenn sie durch ein Programm geleistet werden soll. So fielen zunächst einmal drei Testverfahren weg, die im Programm TESTVERFAHREN enthalten sind: der Wald/Wolfowitz-Runtest, der Dixon/Mood-Vorzeichen-test und der Mediantest. Der Wald/Wolfowitz-Test ist durch den wesentlich effizienteren Kolmogoroff/Smirnow-Anpassungstest oder den χ^2 -Anpassungstest zu ersetzen. Die beiden anderen Verfahren gelten als "Schnelltests" und können durch den Wilcoxon-Test für verbundene Paare bzw. den Mann/Whitney-U-Test ersetzt werden. Für den Restbestand von 16 Testverfahren gilt die Eindeutigkeit der Auswahl in bestimmten Problemsituationen.

Das Programm besteht aus verschiedenen Blöcken, ähnlich wie TESTVERFAHREN: der Initialisierung, den für die Tests benötigten gemeinsamen Unterprogrammen, einigen für die Auswahl benötigten Unterprogrammen, der Auswahl selbst, den Tests und den Fehlerrouitinen. Es beansprucht etwas über 18 Kbyte Speicherplatz, hinzu kommen 11 bis 12 K für die Variablen.

4.3.1 Initialisierung und Fehlerrouitinen

Hier ist ähnliches zu sagen wie bei den entsprechenden Blöcken von TESTVERFAHREN. Die Initialisierung besteht zum einen aus der Dimensionierung der benötigten Arrays, der Festlegung einiger Konstanten und der Definition der Funktionen STN und R. Zum anderen gehören dazu die Zeilen 20 und 30 mit der Aufforderung, die Datendiskette einzulegen, und das Lesen des VARCATALOG. Die Fehlerrouitinen sind dieselben wie beim Programm TESTVERFAHREN.

4.3.2 Unterprogramme

Die Zeilen 100 bis 980 sind identisch mit den entsprechenden Zeilen des Programms TESTVERFAHREN und müssen deshalb an dieser Stelle nicht mehr besprochen werden.

Die Zeilen 1000 bis 1100 enthalten jedoch bisher noch nicht verwendete Unterprogramme, die für die Menus und Abfragen benötigt werden, die die Auswahl des geeigneten Tests ermöglichen.

Die Zeilen 1000 bis 1020 bewirken eine Bildschirm-anzeige, in der nach der Art der zu prüfenden Unterschiede gefragt wird. Eingang der subroutine ist Aß, das vom Programm mit "Verteilungen" bzw. "Mittelwerte" belegt wird. Dieses Unterprogramm wird also für jede Auswahl zweimal durchlaufen. Auf die Frage kann mit J(Ja) oder N(Nein) geantwortet werden. Bei einer anderen Eingabe folgt die schon bekannte Anweisungsfolge zur wiederholten Anforderung einer Eingabe. Bei zugelassener Eingabe wird zum Hauptprogramm zurückgekehrt.

Die Zeilen 1030 bis 1060 liefern eine Meldung, falls ein Test auf Mittelwertunterschiede für abhängige oder unabhängige Stichproben bei nominalem Skalenniveau durchgeführt werden soll. Ein solches Verfahren steht nicht zur Verfügung. Es besteht die Möglichkeit, zu einem Anpassungstest überzugehen. Wird die entsprechende Frage bejaht, erfolgt die Rückkehr zum Hauptprogramm, ansonsten wird zum Beginn der Abfragen (Zeile 2000) gesprungen.

In einigen Problemstitutionen ist für eine Entscheidung die Kenntnis der Anzahl der Kategorien der Untersuchungsvariablen nötig. Die entsprechende Eingabe wird in den Zeilen 1070 bis 1090 angefordert.

Zeile 1100 enthält Hilfen bei der Entscheidung über das vorliegende Meßniveau. Hierzu wird ein Menu angezeigt mit den drei verschiedenen Datenqualitäten. Falls der Anwender nicht sofort angeben

kann, welches Skalenniveau seine Daten haben, kann er das Unterprogramm 1100 aufrufen, daß ihm Erläuterungen zu den verschiedenen Meßniveaus anzeigt.

Damit ist der Bereich der gemeinsamen Unterprogramme abgeschlossen; auf eine Übersicht kann wegen der geringen Anzahl neuer Programmenteile verzichtet werden.

4.3.3 Auswahl eines geeigneten Tests

Die Entscheidung über das zu rechnende Testverfahren wird durch mehrere Menus und Abfragen vorbereitet. Diese erstrecken sich im wesentlichen über die Zeilen 2000 bis 2200. Die Zeilen 2210 bis 2360 enthalten die Verzweigungen zu den Testverfahren. Für diese Sprünge werden mehrere Variablenwerte benötigt. P% bezeichnet die gewählte Option für den Problemtyp, die zwischen 1 und 5 liegen kann. M% steht für das Meßniveau; die Variable hat Werte zwischen 1 und 3. S% bezeichnet die Anzahl der Stichproben; es muß deshalb größer oder gleich zwei sein, da bei Einstichprobentests auf eine Eingabe für S% verzichtet wird.

PS% steht für die genauere Problemspezifikation. Diese Variable wird ebenfalls nur bei den Problemtypen 3 und 4, d. h. bei Prüfung von mehreren unabhängigen oder verbundenen Stichproben erfragt. Sie kann die Werte 1 und 2 annehmen.

R% wird nur in einigen wenigen Situationen angefragt, wenn die Anzahl der Kategorien der Untersuchungsvariablen von Bedeutung für die Entscheidung ist.

Zeile 20000 beginnt mit der Anweisung FRE(0), durch die alle alten Strings aus dem Speicher entfernt werden. Nur die jeweils letzte Belegung einer Stringvariablen bleibt erhalten. Anschließend werden die für die Auswahl benötigten Variablen initialisiert. Es folgt das erste Menu für die Eingabe des Problemtyps (P%) bzw. die Rückkehr zum Hauptmenu:

AUSWAHL MIT HILFEN

WAS SOLL GEPRUEFT WERDEN?

ABWEICHUNG EINER EMPIRISCHEN VER-
TEILUNG VON EINER THEORETISCHEN...1

ZUFAELLIGKEIT EINER STICHPROBE....2

UNTERSCHIEDE ZWISCHEN GRUPPEN.....3

UNTERSCHIEDE ZWISCHEN MESSWERT-
REIHEN DER GLEICHEN GRUPPE.....4

ZUSAMMENHANG ZWISCHEN MERKMALEN...5

REP-TEST MENU.....6

AUSWAHL.....

Wird in Zeile 2030 für P% die Option 6 eingegeben, erscheint, durch Zeile 2040 hervorgerufen, die Aufforderung, die Systemdiskette einzu-

legen, anschließend erfolgt das Laden des Programms REP-TEST MENU, über das der Programmlauf beendet werden kann oder Korrekturen in den Daten vorgenommen werden können. In Zeile 2050 beginnt das nächste Menu über die Datenqualität. Die Bildschirmanzeige sieht zunächst so aus:

WELCHE DATENQUALITAET LIEGT VOR ?

EINGABE:

METRISCHE DATEN.1

ORDINALDATEN....2

NOMINALDATEN....3

HILFE ERHALTEN SIE DURCH DIE H-TASTE

Fehler in der Eingabe werden in Zeile 2070 durch die übliche Anweisungsfolge abgefangen; die Eingabe muß wiederholt werden. Wenn in 2070 die H-Taste gedrückt wird, erfolgt in Zeile 2080 der Sprung ins Unterprogramm 1100, das Erläuterungen zu den drei Optionen über die Datenqualität liefert. Die Bildschirmanzeige wird dadurch erweitert zu folgendem Tableau:

WELCHE DATENQUALITAET LIEGT VOR ?

EINGABE:

METRISCHE DATEN.1	UNTERSCHIEDUNG, RANGORDNUNG UND ABSTAEENDE ERKENNBAR
ORDINALDATEN....2	UNTERSCHIEDUNG UND RANGORDNUNG MOEGL.
NOMINALDATEN....3	NUR UNTERSCHIEDUNG MOEGLICH

HILFE ERHALTEN SIE DURCH DIE H-TASTE

In Zeile 2090 wird dann das Meßniveau (M%) festgelegt. Eingaben außerhalb des Bereichs 1 bis 3 werden zurückgewiesen und es wird erneut die Eingabe angefordert. Für die Problemtypen 1, 2 und 5 (Anpassung an eine theoretische Verteilung, Zufälligkeit einer Stichprobenziehung und Unabhängigkeit von zwei Variablen) werden die Zeilen 210 bis 220 übersprungen, da die Anzahl der Stichproben (S%) und die Problemspezifikation (PS%) nicht erfragt werden müssen.¹⁾

In 210 beginnt nun die Verzweigung zu den verschiedenen Testverfahren.

1) S% ist schon durch den Problemtyp festgelegt.

Falls $P\%=3$ oder $P\%=4$ gewählt wurde, d. h. mehrere unabhängige oder abhängige Stichproben verglichen werden sollen, wird in Zeile 213Ø ihre Anzahl $S\%$ erfragt. Die Zeile 211Ø und 212Ø dienen zur unterschiedlichen Gestaltung der Bildschirmanzeige im einen oder anderen Fall. Alphanumerische Eingaben bzw. eine Stichprobenanzahl kleiner als zwei werden zurückgewiesen durch Zeile 214Ø, Zeile 215Ø enthält die Belegung von $S\%$. Für die beiden Optionen 3 und 4 ist es außerdem notwendig die Problemspezifikation zu erfragen, d. h. ob Unterschiede in den Verteilungen ($PS\%=1$) oder speziell Mittelwertunterschiede ($PS\%=2$) geprüft werden sollen. Die Zeilen 216Ø und 217Ø liefern die Bildschirmanzeigen für die beiden Fälle über das Unterprogramm 1ØØØ und die entsprechende Belegung von $A\%$. Die Frage "Wollen Sie Unterschiede in den ... prüfen?" kann mit J(Ja) oder N(Nein) beantwortet werden. Falls eine der beiden Fragen bejaht wurde, wird $PS\%$ entsprechend belegt und der Beginn der Verteilung in Zeile 221Ø angesprungen. Sind hingegen beide Fragen verneint worden, erfolgt durch Zeile 218Ø ein Hinweis, daß nicht mehr Möglichkeiten für die Problemspezifikationen zur Verfügung stehen. Durch Drücken der RETURN-Taste kann zum ersten Menu in Zeile 2ØØØ zurückgekehrt werden. Wird eine andere Taste gedrückt, werden die beiden Möglichkeiten der Problemspezifikation noch einmal angezeigt und die Eingabe kann bei Bedarf korrigiert werden, d. h. die Zeilen ab 216Ø werden erneut durchlaufen.

Zeile 2210 enthält den Verteiler für die verschiedenen Problemtypen.

Für $P\%=1$ (Anpassung an eine theoretische Verteilung) werden die Zeilen 2220 bis 2240 bearbeitet. Für $M\%=1$, d. h. für metrisches Meßniveau, wird in Zeile 13000 gesprungen, nämlich zum Kolmogoroff/Smirnow-Einstichproben-test. Ansonsten muß eine Eingabe für $R\%$ (die Anzahl der Kategorien der Untersuchungsvariablen) verlangt werden. Dies geschieht in Zeile 2230 über das Unterprogramm 1070. Nach der Belegung der Variablen $R\%$ folgt bei ordinalem Meßniveau der Sprung zum Kolmogoroff/Smirnow-Test, falls $R\%<5$ ist, und zum χ^2 -Anpassungstest in Zeile 12000, wenn $R\%\geq 5$ ist. Bei nominalem Meßniveau erfolgt die Verzweigung zum Binomialtest in Zeile 11000, wenn $R\%=2$ ist, bei größeren Werten von $R\%$ zum χ^2 -Anpassungstest. Kleinere Werte von $R\%$ als zwei sind im Unterprogramm 1070 als fehlerhafte Eingaben zurückgewiesen worden.

Für $P\%=2$ werden die Zeilen 2250 und 2260 durchlaufen. Die Zufälligkeit einer Stichprobenziehung wird für ordinales und metrisches Skalenniveau ($M\%<3$) über den Runtest in Zeile 14000 geprüft. Für nominales Meßniveau ist eine Dichotomisierung am Median oder am arithmetischen Mittel nicht möglich, es kann deshalb kein Test durchgeführt werden. Eine entsprechende Meldung erfolgt durch Zeile 2260.

Für $P\%=3$, also die Prüfung unabhängiger Stichproben werden die Zeilen 2270 bis 2320 benötigt.

Bei Problemspezifikation 1, also Prüfung auf Anpassung der Verteilungen wird in den Zeilen 2280 bis 2300 ein geeignetes Verfahren ausgewählt. Für mehr als zwei Stichproben ($S \geq 2$) wird unabhängig vom Meßniveau ein χ^2 -Homogenitätstest (Zeile 15000) durchgeführt. Für zwei Stichproben wird bei metrischem oder ordinalem Meßniveau ($M \leq 3$) ein Kolmogoroff/Smirnow-Zweistichprobentest gerechnet, der sich ab Zeile 16000 befindet. Für nominales Meßniveau ist die Entscheidung abhängig von der Anzahl der Kategorien der Untersuchungsvariablen.

Über das Unterprogramm 1070 wird diese Zahl R erfragt. Bei dichotomer Variable ($R=2$) wird ein Fisher-Test (Zeile 17000) gerechnet. Für eine polytome Variable erfolgt der Sprung in Zeile 15000 zum χ^2 -Homogenitätstest. Über den Stichprobenumfang braucht zunächst nichts bekannt zu sein. Falls er zu groß ist für die exakte Berechnung über den Fisher-Test, wird ohnehin durch das Programm zu einem χ^2 -Homogenitätstest übergegangen.

Falls Mittelwertunterschiede geprüft werden sollen, also $PS=2$ ist, sind die Zeilen 2310 und 2320 relevant. Für Nominaldaten ($M=3$) steht kein Test auf Mittelwertunterschiede zur Verfügung. Durch das Unterprogramm in Zeile 1030 erfolgt eine entsprechende Meldung und die Frage, ob stattdessen ein Anpassungstest durchgeführt werden soll. Falls die Frage bejaht wurde, wird in Zeile 2280 gesprungen, um zu klären, welcher Anpassungstest in Frage kommt. Bei Verneinung erfolgt schon durch das Unterprogramm die Rückkehr zum Menu in

Zeile 2000. Für höhere Datenqualität wird je nach Anzahl der Stichproben ein Mann/Whitney-U-Test ($S=2$) über die Zeilen 18000 ff. oder ein Kruskal/Wallis-H-Test ($S>2$) über die Zeilen 19000 ff. durchgeführt.

Die Zeilen 2330 bis 2350 werden abgearbeitet wenn $P=4$ ist, d. h. verbundene Stichproben vorliegen. Als Anpassungstests stehen hier die Verfahren von McNemar (Zeile 20000) und Cochran (Zeile 21000) zur Verfügung, die je nach Stichprobenanzahl S eingesetzt werden (Zeile 2330).

Für nominales Meßniveau ist wiederum kein Test auf Mittelwertunterschiede im System REP-TEST enthalten. Die Zeile 2340 bewirkt - analog zur Zeile 2310 bei $P=3$ - eine entsprechende Meldung. Auf Wunsch kann stattdessen ein Anpassungstest gewählt werden.

Auf höherem Skalenniveau ist die Entscheidung von der Anzahl der Stichproben abhängig.

Möglich sind der Wilcoxon-Vorzeichen-Rangtest (ab Zeile 22000) für $S=2$ und der Friedman-Test (ab Zeile 23000) für $S>2$.

Die letzte Zeile vor Beginn der Testverfahren, Zeile 2360, leistet die Verzweigung zu den Testverfahren für Problemtyp 5, den Zusammenhang zwischen Merkmalen. Je nach Meßniveau wird der Korrelationskoeffizient von Bravais/Pearson (ab Zeile 24000) oder der Rangkorrelationskoeffizient nach Spearman (Zeile 25000 ff.) geprüft bzw. ein χ^2 -Unabhängigkeitstest durchgeführt (Sprung zur Zeile 26000).

Damit ist für alle Situationen die Auswahl des geeigneten Verfahrens getroffen worden. Es soll noch darauf hingewiesen werden, daß für diesen wegen der vielen Verzweigungen etwas unübersichtlichen Teil des Programms darauf verzichtet wurde, die Zeilen möglichst zusammenzufassen. Um die Lesbarkeit des Programms nicht vollständig auszuschließen, wurde mehr auf eine sinngemäße Gliederung geachtet. Da im Gegensatz zum Programm TESTVERFAHREN nur 16 Tests aufgenommen werden mußten, konnte der Platz für den Bereich Menus und Abfragen etwas verschwenderisch bemessen werden. Einen Überblick über die Verzweigungen der Zeilen 221Ø bis 236Ø kann man sich durch das folgende Schaubild verschaffen.

$P_{\%} = 1$	$M_{\%} = 1$	→	13000	(K-S-Test)
	$M_{\%} = 2$	→	12000	(χ^2 -Test)
	$R_{\%} < 5$	→	13000	(K-S-Test)
	$R_{\%} > 5$	→	11000	(Binomial test)
	$M_{\%} = 3$	→	12000	(χ^2 -Test)
$P_{\%} = 2$	$M_{\%} = 1, 2$	→	14000	(Runttest)
	$M_{\%} = 3$	→	2000	(Menu)
$P_{\%} = 3$	$PS_{\%} = 1$	→	16000	(K-S-Test)
	$S_{\%} = 2$	→	17000	(Fisher-Test)
	$M_{\%} = 3$	→	15000	(χ^2 -Test)
	$S_{\%} > 2$	→	15000	(χ^2 -Test)
	$M_{\%} = 1, 2$	→	18000	(Mann/Whitney)
	$S_{\%} > 2$	→	19000	(Kruskal/Wallis)
	$M_{\%} = 3$	→	2000	(Menu)
	$\{ \text{Meldung: kein Test verfügbar} \}$	→	PS $_{\%}$ =1	
	$\{ \text{Anpassungstest?} \}$	→		
	n	→		
$P_{\%} = 4$	$PS_{\%} = 1$	→	20000	(McNemar)
	$S_{\%} > 2$	→	21000	(Cochran)
	$M_{\%} = 1, 2$	→	22000	(Wilcoxon)
	$S_{\%} > 2$	→	23000	(Friedman)
	$M_{\%} = 3$	→	2000	(Menu)
$P_{\%} = 5$	$PS_{\%} = 2$	→	PS $_{\%}$ =1	
	$\{ \text{Meldung: kein Test verfügbar} \}$	→		
	$\{ \text{Anpassungstest?} \}$	→		
	n	→		
$P_{\%} = 5$	$M_{\%} = 1$	→	24000	(Bravais/Pearson)
	$M_{\%} = 2$	→	25000	(Spearman)
	$M_{\%} = 3$	→	26000	(χ^2 -Unabh.)

4.3.4 Testverfahren

Zu diesem Punkt ist im Abschnitt 4.2 soviel gesagt worden, daß hier darauf verzichtet werden kann. Es sollen nur die enthaltenen Tests mit den zugehörigen Zeilennummern aufgelistet werden. Gelegentlich sind einige Anmerkungen notwendig.

Zeilen	Test
11000-11060	Binomialtest
12000-12090	χ^2 -Anpassungstest
13000-13350	Kolmogoroff/Smirnow-Einstichproben- test
14000-14080	Runtest Es ist die Zeile 14010 entfallen, da vom Programm für metrisches Skalenniveau die Dichotomisierung am arithmetischen Mittel und für ordinales Meßniveau die Dichotomisierung am Median vorgenommen wird.
15000-15080	χ^2 -Homogenitätstest
16000-16040	Kolmogoroff/Smirnow-Zweistichproben- test
17000-17020	Fisher-Test
18000-18040	Mann/Whitney-U-Test

Zeilen	Test
19000-19040	Kruskal/Wallis-H-Test
20000-20100	McNemar-Test Es wurde eine Umnummerierung vorgenommen, der Ablauf blieb unverändert.
21000-21060	Cochran-Q-Test Da die Anzahl der Stichproben schon im 2000er-Block erfragt wurde, kann hier auf eine Eingabe für VC% verzichtet werden. Dadurch ändert sich die Zeile 21000.
22000-22070	Wilcoxon-Test für verbundene Paare
23000-23050	Friedman Die Zeile 23000 ist leicht verändert, da eine Eingabe für VC% nicht erforderlich war. Es konnte VC%=S% gesetzt werden.
24000-24030	Bravais/Pearson
25000-25050	Spearman
26000-26100	χ^2 -Unabhängigkeitstest

Kapitel 5 Beispiele

Für dieses Kapitel wurde der Datenbestand einer empirischen Untersuchung zur Situation ausländischer Studenten an der Ruhr-Universität Bochum verwendet. Die Umfrage wurde von der "Arbeitsgruppe Ausländerfragen des Universitätsparlaments der Ruhr-Universität Bochum" durchgeführt; es handelte sich um eine schriftliche Befragung bei allen 1000 ausländischen Kommilitonen der RUB. Da die Rücklaufquote äußerst gering war (7,5 %), konnte die Arbeitsgruppe nur eine deskriptive Auswertung der Daten vornehmen. Der Einsatz inferenzstatistischer Methoden war unter diesen Umständen nicht vertretbar. Aus diesem Grund wurde in der vorliegenden Arbeit darauf verzichtet, einen theoretischen Rahmen für die Diskussion der soziologischen Grundlagen von Integration von Ausländern und ihrer Hindernisse abzustecken. Stattdessen wird das Datenmaterial benutzt, um einige Tests beispielhaft vorzurechnen und die Ergebnisausdrucke vorzustellen und zu erläutern. Es muß jedoch eindringlich darauf hingewiesen werden, daß Rückschlüsse auf die Realität ausländischer Studenten oder allgemein ausländischer Mitbürger nicht möglich sind. Da von den 1000 verteilten Fragebogen nur 75 wieder abgegeben wurden, ist klar, daß von einer Interpretation der Testergebnisse über ihre rein formale Aussage hinaus abgesehen werden muß.

So wurde denn darauf verzichtet, Fragebogen und Codeplan der Untersuchung eingehend zu erläutern. Zum Verständnis der durchgeführten Tests wurden sie lediglich in den Anhang mit aufgenommen.

Nicht alle Testverfahren lassen sich auf die vorliegenden Daten anwenden. Wenn verbundene Stichproben vorliegen, so heißt das, dieselben Untersuchungselemente werden zu mehreren Zeitpunkten, unter verschiedenen Bedingungen o. ä. auf die Ausprägungen einer Untersuchungsvariablen untersucht. Statt einer Stichprobe können auch mehrere homogenisierte Gruppen als Grundlage der Erhebung dienen. Dies war nun anhand des gegebenen Datenmaterials nicht vorzuführen. Um zumindest die Ergebnisdrukke erstellen zu können, wurden für diese Fälle mehrere Untersuchungsvariablen anstelle einer Variablen unter mehreren Bedingungen gewählt. Nun zu den Beispielen.

- 1) Zweimal wurde die Verteilung der Variablen Geschlecht, bezeichnet mit V2, auf ihre Anpassung an eine Binomialverteilung überprüft. Zunächst über die Voreinstellung des Tests, d. h. mit dem Bernoulli-Parameter 0,5, danach mit $\pi = 0,2$.

Die Stichprobenrealisation ist bei 0,5 signifikant - $S = 0$ -, bei 0,2 hingegen erst für Signifikanzniveaus größer oder gleich 35 %, wie sie in der Praxis wohl nicht gewählt werden.

REP-TEST

-----BINOMIAL-----

VARIABLE V2

CODE	HAEUF.	REL.	WAHRSCH.
0	55	.775	1
1	16	.225	0

BERNOULLI-PARAMETER .5

GUELTIGE FAELLE 71/ FEHLENDE FAELLE 4

REP-TEST

-----BINOMIAL-----

VARIABLE V2

CODE	HAEUF.	REL.	WAHRSCH.
0	55	.775	.65
1	16	.225	.35

BERNOULLI-PARAMETER .2

GUELTIGE FAELLE 71/ FEHLENDE FAELLE 4

- 2) Ein χ^2 -Anpassungstest wurde für die Variable V5, das 1. Studienfach der Befragten, durchgeführt. Es sind fünf verschiedene Ausprägungen möglich.¹⁾ Bei der Prüfung auf Gleichverteilung erwies sich die Stichprobenrealisation als signifikant.

REP-TEST

-----CHI QU.-ANPASSUNGSTEST-----

VARIABLE V5

CODE	BEOBACHTET	ERWARTET
0	28	13.4
1	8	13.4
2	18	13.4
3	7	13.4
4	6	13.4
FREIHEITSGRADE		4
TESTWERT CH		26.806
SIGNIFIKANZ		0

GUELTIGE FAELLE 67/ FELENDE FAELLE 8

Ein weiterer χ^2 -Anpassungstest wurde für die Variable V4, das Herkunftsland, gerechnet. Die Codierung erfolgte in 7 Kategorien. Für dieses Beispiel wurden die erwarteten Häufigkeiten nicht durch das Programm berechnet, sondern vom Benutzer eingegeben, wie es der

1) Zu den Codierungen siehe den Codeplan im Anhang

Tabelle zu entnehmen ist. Der Testwert von 14,027 ist bei 6 Freiheitsgraden signifikant für $\alpha \geq 0,029$.

REP-TEST

-----CHI QU.-ANPASSUNGSTEST-----

VARIABLE V4

CODE	BEOBACHTET	ERWARTET
0	1	5
1	8	5
2	8	5
3	9	5
4	8	5
5	18	25
6	13	15

FREIHEITSGRADE	6
TESTWERT CH	14.027
SIGNIFIKANZ	.029

GUELTIGE FAELLE 65/ FEHLENDE FAELLE 10

- 3) Der Kolmogoroff/Smirnow-Einstichprobentest wurde in zwei Versionen gerechnet. Einmal zur Prüfung der Variable Alter (V3) auf Anpassung an die Normalverteilung. Das Ergebnis ist bis zu $\alpha = 0,06$ nicht signifikant. Zum anderen wurde ein Test auf Anpassung an die Poissonverteilung für die Variable V9, die Aufenthaltsdauer in der Bundesrepublik Deutschland in Jahren, durchgeführt, der zu

einem signifikanten Ergebnis führte.

REP-TEST

-KOLMOGOROFF/SMIRNOW-1 STICHPROBEN-TEST-

VARIABLE U3 NORMALVERT.

	ABS. DIFF.	POS. DIFF.	NEG. DIFF.
MAX.	.161	.161	-.075

TESTWERT Z 1.326

SIGNIFIKANZ .06

GUELTIGE FAELLE 68/ FEHLENDE FAELLE 7

REP-TEST

-KOLMOGOROFF/SMIRNOW-1 STICHPROBEN-TEST-

VARIABLE U9 POISSONVERT.

	ABS. DIFF.	POS. DIFF.	NEG. DIFF.
MAX.	.286	-	-.286

TESTWERT Z 2.362

SIGNIFIKANZ 0

GUELTIGE FAELLE 68/ FEHLENDE FAELLE 7

- 4) Mit dem χ^2 -Homogenitätstest wurde überprüft, ob sich die vier Gruppen, die durch unterschiedliche Semesterzahlen entstehen (das Merkmal wurde in klassifizierter Form erfaßt), hinsichtlich der Verteilung der Variable Herkunftsland wesentlich unterscheiden. Das Beispiel wurde gewählt, um einen Test mit relativ hoher Zahl von Freiheitsgraden vorführen zu können.

REP-TEST

-----CHI QU.-HOMOGENITAETSTEST-----

VARIABLE V4 MIT VARIABLE V8

FREIHEITSGRADE	18
TESTWERT CHI	25.188
SIGNIFIKANZ	.12
GUELTIGE FAELLE 62/ FEHLENDE FAELLE 13	

Für das zweite Beispiel wurden zwei dichotome Merkmale gewählt. Zur Gruppierung diente die Variable V7 (es handelt sich um die Frage, ob vor dem Aufenthalt in der Bundesrepublik schon ein Studium begonnen worden ist). Die beiden Stichproben wurden auf Unterschiede der Verteilung der Variable Geschlecht untersucht.

REP-TEST

-----CHI QU.-HOMOGENITAETSTEST-----

VARIABLE V2 MIT VARIABLE V7

FREIHEITSGRADE	1
TESTWERT CHI	2.784
SIGNIFIKANZ	.095
GUELTIGE FAELE 69/ FEHLENDE FAELE 6	

- 5) Für den Kolmogoroff/Smirnow-Zweistichproben-test wird nur ein Beispiel vorgestellt. Die Frage nach dem Wunsch zur Verbesserung der Sprachkenntnisse (V58) kann mit ja oder nein beantwortet werden.

Die Ausprägungen dieser dichotomen Variablen gruppieren die Stichproben. Geprüft wird die Verteilung der Variablen V64, die Frage nach dem Wohlfühlen in Deutschland. Sie ist ordinalskaliert mit 5 Stufen. Die Verteilung in den beiden Stichproben ist gleich, die Signifikanz beträgt 1.

```
***REP-TEST***  
-KOLMOGOROFF/SMIRNOW-2 STICHPROBEN-TEST-  
VARIABLE V64 MIT VARIABLE V58  
CODE 0: 15 CODE 1: 48  
ABS. DIFF. POS. DIFF. NEG. DIFF.  
MAX. .096 .067 -.096  
  
TESTWERT Z .324  
SIGNIFIKANZ 1  
GUELTIGE FAELLE 63/ FEHLENDE FAELLE 12
```

- 6) Der Wald/Wolfowitz-Runtest wird vorgeführt am Beispiel der Variable V8 (Semesterzahl), deren Werte durch das Geschlecht (V2) gruppiert werden. Es ist kein Unterschied zwischen den Verteilungen der beiden Stichproben festzustellen, S ist gleich 0,413. Da aber Sign-Wechsel innerhalb von Verbundwertgruppen aufgetreten sind, ist das Ergebnis vorsichtig zu interpretieren.

REP-TEST

-----WALD/WOLFWITZ-RUNTEST-----

VARIABLE V8 MIT VARIABLE V2

CODE 0 53 CODE 1 15

ANZ. DER RUNS 25

TESTWERT Z .221

SIGNIFIKANZ .413

GUELTIGE FAELLE 68/ FEHLENDE FAELLE 7

ACHTUNG

ES TRETEN SIGN-WECHSEL AUF
INNERHALB VON TIE-GRUPPEN.

- 7) Der McNemar-Test für verbundene Stichproben wird nur mit seinem Ergebnisausdruck vorgestellt:

REP-TEST

-----MCNEMAR-----

VARIABLE V2 MIT VARIABLE V56

		V2		
		0	1	
V56		I-----I-----I		
	0	I 22	I 9	I
		I-----I-----I		
	1	I 28	I 6	I
		I-----I-----I		

TESTWERT CH 8.757

SIGNIFIKANZ 3E-03

GUELTIGE FAELLE 65/ FEHLENDE FAELLE 10

- 8) Das Ergebnis des Runtests für die Variable Alter (V3) widerspricht nicht der Hypothese der Zufälligkeit der Stichprobe.

```
***REP-TEST***  
-----RUNS-----  
  
VARIABLE V3  
  
ANZ.DER RUNS                38  
ANZ. DER + DIFFERENZEN     40  
ANZ. DER - DIFFERENZEN     28  
TESTWERT Z                  1.024  
SIGNIFIKANZ                 .306  
  
GUELTIGE FAELLE 68/ FEHLENDE FAELLE 7
```

- 9) Für den χ^2 -Unabhängigkeitstest werden zwei Ergebnisprotokolle erstellt, zum einen das Ergebnis des Tests, zum anderen auf Wunsch eine bivariate Häufigkeitstabelle. Zunächst werden die beiden Variablen Geschlecht (V2) und Studienfach (V5) auf Unabhängigkeit geprüft. Das Ergebnis widerspricht bei relativ niedrigem Signifikanzniveau der Hypothese.

REF-TEST

-----CHI QU.-UNABHAENGIGKEITSTEST-----

VARIABLE V2 MIT VARIABLEV5

FREIHEITSGRADE	4
TESTWERT CH	15.969
SIGNIFIKANZ	3E-03

GUELTIGE FAELLE 67/ FEHLENDE FAELLE 8

REF-TEST

-----CHI QU.-UNABHAENGIGKEITSTEST-----

V5

CODE	0	1	2	3	4
------	---	---	---	---	---

V2

CODE 0	28	5	10	4	5
CODE 1	0	3	8	3	1

GUELTIGE FAELLE 67/ FEHLENDE FAELLE 8

Das zweite Beispiel illustriert den Fall, daß die Prüfgröße so groß wird, daß die Signifikanz nicht mehr berechnet werden kann. Es handelt sich um die Prüfung der Unabhängigkeit bei den beiden Variablen V5 und V11 (angestrebter Studienabschluß). CH ist so groß, daß S auf 999 gesetzt wird.

REF-TEST

-----CHI QU.-UNABHAENGIGKEITSTEST-----

VARIABLE V11 MIT VARIABLE V5

FREIHEITSGRADE	16
TESTWERT CH	110.121
SIGNIFIKANZ	999

GUELTIGE FAELLE 64/ FEHLENDE FAELLE 11

REF-TEST

-----CHI QU.-UNABHAENGIGKEITSTEST-----

V5

CODE	0	1	2	3	4
------	---	---	---	---	---

V11

CODE 0	25	0	7	5	0
CODE 1	0	2	4	0	0
CODE 2	3	1	3	2	0
CODE 3	0	0	1	0	6
CODE 4	0	5	0	0	0

GUELTIGE FAELLE 64/ FEHLENDE FAELLE 11

- 10) Der Test für den Rangkorrelationskoeffizienten von Spearman wurde angewandt auf die beiden ordinalskalierten Merkmale V64 (Wohlfühlen in Deutschland) und V66 (Eigene Einschätzung der Deutschkenntnisse). Das Testergebnis läßt einen Zusammenhang zwischen den beiden Variablen ver-

muten. Allerdings keinen sehr starken, denn r ist nur gleich 0,368.

REP-TEST

-----SPEARMAN-----

VARIABLE V64 MIT VARIABLE V66

	VARIABLE1	VARIABLE2
ANZ. DER TIES	64	63
KORRELATIONSKOEFF.		.368
FREIHEITSGRADE		62
TESTWERT T		3.121
SIGNIFIKANZ		3E-03
GUELTIGE FAELE 64/ FELENDE FAELE 11		

- 11) Für den Korrelationskoeffizienten von Bravais/Pearson gab es nur eine Möglichkeit für ein Beispiel: Alter und Aufenthaltsdauer. Das Ergebnis läßt auf Unabhängigkeit schließen.

REP-TEST

-----BRAVAIS/PEARSON-----

VARIABLE V3 MIT VARIABLE V9

KORRELATIONSKOEFF. R	.046
FREIHEITSGRADE	63
TESTWERT T	.367
SIGNIFIKANZ	.715

GUELTIGE FAELLE 65/ FEHLENDE FAELLE 10

- 12) Der Mediantest wird für eine dichotome Gruppierungsvariable vorgeführt. Das Ergebnis läßt die Ablehnung der Hypothese auf Gleichheit der beiden Stichprobenmediane in den Gruppen der männlichen und der weiblichen Kommilitonen für die Variable Semesterzahl nicht zu.

REP-TEST

-----MEDIANTEST-----

VARIABLE V8 MIT VARIABLE V2

MEDIAN: 1		'<='	'>'	
	CODE 0	29	24	53
	CODE 1	10	5	15
		39	29	68

FREIHEITSGRADE	1
TESTWERT CHI	.281
SIGNIFIKANZ	.596

GUELTIGE FAELLE 68/ FEHLENDE FAELLE 7

- 13) Für den Mann/Whitney-U-Test wurde zum Vergleich dasselbe Beispiel gewählt.

```

***REP-TEST***

-----MANN/WHITNEY-U-TEST-----

VARIABLE V8 MIT VARIABLE V2

CODE  ANZAHL  MITTL. RANG

    0      53      35.613
    1      15      30.567

U      338.5      ANZ. DER TIES      68

TESTWERT Z                      -.915

SIGNIFIKANZ                      .36

GUELTIGE FAELLE 68/ FELENDE FAELLE 7

```

- 14) Für den Kruskal/Wallis-H-Test wurde als Untersuchungsvariable das metrische Merkmal Alter gewählt, die Gruppierung erfolgt durch V11 (angestrebter Studienabschluß), die 5 Kategorien hat.

REP-TEST

-----KRUSKAL/WALLIS-H-TEST-----

VARIABLE V3 MIT VARIABLE V11

CODE ANZAHL MITTL. RANG

0	36	28.711
1	5	34.8
2	9	46.833
3	7	23.786
4	5	45.4

FREIHEITSGRADE 4 TIES: 59

TESTWERT CHI 11

SIGNIFIKANZ .027

GUELTIGE FAELLE 64/ FEHLENDE FAELLE 11

- 15) Der Ergebnisausdruck des Dixon/Mood-Vorzeichen-tests wird ebenso wie die der beiden folgenden Tests für verbundene Stichproben ohne Erläuterung vorgestellt. Die Testvariablen stehen für verbundene Meßwertreihen.

REP-TEST

-----DIXON/MOOD-VORZEICHEN-TEST-----

VARIABLE V63 MIT VARIABLE V64

ANZ. DER + DIFFERENZEN 22

ANZ. DER - DIFFERENZEN 10

TESTWERT Z 1.945

SIGNIFIKANZ .052

GUELTIGE FAELLE 65/ FEHLENDE FAELLE 10

REP-TEST

--WILCOXON-TEST FUER VERBUNDENE PAARE---

VARIABLE V63 MIT VARIABLE V64

ANZ. DER TIES	32
+ DIFFERENZEN 22	MITTL. RANG 17.773
- DIFFERENZEN 10	MITTL. RANG 13.7
TESTWERT Z	-2.375
SIGNIFIKANZ	. .018

GUELTIGE FAELLE 65/ FEHLENDE FAELLE 10

REP-TEST

-----FRIEDMAN-----

TESTVARIABLEN	MITTLERE RAENGE
V61	1.755
V65	3.5
V60	2.557
V66	4.179
V64	3.009

FREIHEITSGRADE	4
TESTWERT CHI	71.826
SIGNIFIKANZ	0

GUELTIGE FAELLE 53/ FEHLENDE FAELLE 22

Wenn auch bedauerlicherweise auf weitergehende inhaltliche Ausführungen verzichtet werden mußte, so kann man sich anhand der vorgeführten Beispiele doch ein Bild über die Leistungsfähigkeit der Programme des Systems REP-TEST machen.

Kapitel 6 Schlußbemerkungen

Diese Arbeit hatte das Ziel, den Bereich der statistischen Signifikanztestverfahren neu zu strukturieren. Auf der Grundlage einer anwendungsbezogenen Analyse der Einsatzbedingungen der verschiedenen Verfahren wurden Computerprogramme entwickelt, die den für Sozialwissenschaftler besonders interessanten Bereich der nicht-parametrischen Testverfahren für die Kleinrechneranwender erschließen. Zusätzlich ergab sich durch die neue Gliederung die Möglichkeit, für ungeübte Benutzer statistischer Methoden eine Reihe von Fehlerquellen auszuschließen durch die Entwicklung des Programms TESTS MIT HILFEN, das nicht nur Testverfahren durchführen kann, sondern außerdem auch das geeignete Testverfahren für eine konkrete Anwendungssituation auswählt. Die in Kapitel 3 erwähnten Beschränkungen, die der Benutzer bei der Anwendung eines speziellen Tests beachten sollte, sind in den Programmen zum größten Teil berücksichtigt worden. Dabei wurde dem erfahrenen Statistiker natürlich die freie Entscheidung über das geeignete Verfahren überlassen, d. h. im Programm TESTVERFAHREN konnte auf die Überprüfung von Voraussetzungen, soweit sie nicht die Kapazität des Rechners berührten, weitgehend verzichtet werden.

Das Programm TESTS MIT HILFEN hingegen kontrolliert auch einige Einsatzbedingungen. Selten auftretende Datenkonstellationen, die den Einsatz eines üblichen Tests in Frage stellen, konnten jedoch nicht

in den Auswahlkriterien mit berücksichtigt werden. Um auch in solchen Fällen Fehler auszuschließen, ist ein Computerprogramm mit standardisierten Eingaben ungeeignet. Ein "Instruktionsprogramm" als Nachhilfe für Statistik-Laien war jedoch nicht Ziel dieser Arbeit. Genau das wäre aber nötig, um alle Spezialfälle zu erfassen.

Mit der vorliegenden Arbeit ist es also gelungen, die Einsatzmöglichkeiten von Kleinrechnern um einen zusätzlichen Bereich zu erweitern. Zwar lagen schon Programmpakete für Kleinrechner vor, die auch nicht-parametrische Testverfahren enthielten, doch keine so umfangreiche Zusammenstellung. Micro-SPSS von ROSE (1982) z. B. enthält nur 8 der 19 in REP-TEST enthaltenen Verfahren und ist zudem eine Adaption von SPSS für Kleinrechner - wie schon der Name sagt -, d. h. die statistischen Auswertungsprozeduren werden nur im batch-Betrieb durchgeführt, eine unsinnige Übernahme der Gepflogenheiten beim Großrechnerbetrieb. Zwar kann dieses System bis zu 700 Fälle und bis zu 100 Variablen verarbeiten,¹⁾ doch liegt die Stärke von Kleinrechnern gerade nicht im Bereich der großen Datenbestände. Ein interaktives System wie REP-TEST kann genau da eingesetzt werden, wo der Kleinrechnereinsatz seine Berechtigung hat: bei kleinen bis mittleren Fallzahlen für schnelle Auswertungen in ständiger Kommu-

1) Dies richtet sich auch nach der verfügbaren Hardware-Konfiguration. In jedem Fall werden mindestens 2 Diskettenlaufwerke verlangt.

nikation mit dem Rechner. Auf die komplizierte Syntax der Anweisungen von SPSS-ähnlichen Programmpaketen sollte verzichtet werden, zugunsten eines einfach und relativ schnell zu erlangenden Ergebnisses.

Das System REP-TEST bietet sich für einen speziellen Bereich der Statistik als Ersatz und Ergänzung zum Großrechnereinsatz an, da es ein reichhaltiges Angebot an nicht-parametrischen Tests zur Verfügung hat. Insbesondere wird einem bestimmten Personenkreis, der vielleicht eher Zugang zu einem Kleinrechner als zu einem Großrechner hat, die Durchführung inferenzstatistischer Verfahren erleichtert. Dies war möglich durch anwendungsorientierte Strukturierung des Problemkreises und entsprechenden Aufbau der Programme.

Wenn die neue Technologie der Kleinrechner sinnvoll genutzt werden soll, müssen ihre Stärken und Schwächen erkannt werden. Nur dann können zweckmäßige Programme entwickelt werden und nur dann können die Vorteile der Mikrocomputer wirklich ausgenutzt werden.

Literaturverzeichnis:

- APPLE COMPUTER INC. (Hrsg.): BASIC-Programmierhandbuch, Cupertino 1978
- APPLE COMPUTER INC. (Hrsg.): APPLE II-Benutzerhandbuch, Cupertino 1979
- BAMBERG, G.: Statistische Entscheidungstheorie, Würzburg/Wien 1972
- BAUER, H.: Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie, Berlin 1968
- BEUTEL, P. u.a.: SPSS 7, 2.Aufl., Stuttgart/New York 1978
- BHATTACHARJEE, G.P.: Algorithm AS 32 : The Incomplete Gamma-Integral, in: Applied Statistics 19/1979, S. 285-287
- BORCHERS, L. u. M. POOLE: Some Common BASIC Programs, Berkeley 1979
- BORTZ, J.: Lehrbuch der Statistik für Sozialwissenschaftler, Berlin 1977
- BRANDT, S.: Datenanalyse, Mannheim 1981
- BÜNING, H. u. G. TRENKLER: Nichtparametrische statistische Methoden, Berlin/New York 1978
- CHAMBERS, J.: Computational Methods for Data Analysis, New York 1980
- CHOCHRAN, W.G.: The Comparision of Percentages in Matches Samples, in: Biometrika, 37/1950, S.256-266
- COOVERT, M.D.: An Elementary Statistical Package for 6502-based Microcomputers, in: BRMI 12/1980, S.562
- COX, D.R.: Planning of Experiments, New York 1958
- DAHL, O.J., DIJKSTRA, E.W. u. C.A.R. HOARE: Structural Programming, London/New York 1972
- DAVID, H.A., HARTLEY, H.O. u. E.S. PEARSON: The Distribution of the Ratio in a Single Normal Sample, of Range to Standard Deviation, in: Biometrika 41/1954, S.482-493

- DERENZO, S.E.: Approximations for Hand Calculators Using Small Integer Coefficients, in: Mathematics of Computation, 31/1977, S.214-225
- DIXON, W.J.: Ratio Involving Extreme Values, in: Annals of Mathematical Statistics, 22/1951, S.68-78
- DIXON, W.J. u. A.M. MOOD: The Statistical Sign Test, in: Journal of the American Statistical Association, 41/1946, S.557-566
- DIXON, W.J. u. M.B. BROWN: BMDP 79, Berkeley 1979
- FEGAN, R.L. u. S.L. BROSCHE: BASIC Computer Programming for Statistics, Englewood Cliffs, o.J.
- FERGUSON, T.S.: Mathematical Statistics. A Decision Theoretic Approach, New York/London 1967
- FISHER, R.A.: The Conditions under which Chi-square Measures the Discrepancy between Observation and Hypothesis, in: Journal of the Royal Statistical Society, 87/1924, S.442-450
- FISHER, R.A.: On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample, in: Metron 1, 4/1921, S.1ff
- FISHER, R.A.: Statistical Methods for Research Workers, Edinburgh/London 1941
- FISZ, M.: Wahrscheinlichkeitsrechnung und mathematische Statistik, 7. Aufl., Berlin 1973
- FRANCIS, I.: Criteria and Consideration in the Evaluation of Statistical Program Packages, in: American Stat. 2/1975, S.52-56
- FRANCIS, I.: A Comparative Review of Statistical Software, Voorburg 1979
- FRASER, D.A.S.: Nonparametric Methods in Statistics, New York 1957
- FREEMAN, P.R.: Algorithm AS 26 : Ranking an Array of Numbers, in: Applied Statistics, 19/1974, S.111-113
- FRIEDMAN, M.: A Comparison of Alternative Tests of Significance for the Problem of m rankings, in: Annals of Mathematical Statistics 11/1940, S.86-92

- FRIEDMAN, M.: The Use of Ranks to Avoid the Assumption of Normality Implicit the Analysis of Variance, in: Journal of the American Statistical Association, 32/1937, S.675-701
- GIBBONS, J.: Non-Parametric Methods for Quantitative Analysis, New York 1976
- GILBERT, L.H.: PSYCHOSTATS. BASIC Programs for Data Analysis in Psychology, in: BRMI 11/1979, S.464
- GOLDKUHLE, P.: Direkte Funktionseingabe in BASIC Programmen, in: APPLE USER MAGAZINE 6/1981
- GRUBBS, F.E. u. G. BECK: Extension of Sample Sizes and Percentage Point for Significance Tests of Outlying Observations, in: Technometrics, 14/1972, S.847-854
- GRUBBS, F.E.: Errors of Measurement, Precision, Accuracy and the Statistical Comparison of Measuring Instruments, in: Technometrics, 15/1973, S.53-66
- HÁJEK, J. u. Z. ŠIDÁK: Theory of Rank Tests, Prag/New York 1967
- HART, J.F. (Hrsg.): Computer Approximations, New York 1968
- HEMMERLE, W.J.: Statistical Computations on a Digital Computer, London 1967
- HOROWITZ, E. u. S. SAHNI: Fundamentals of Computer Algorithms, London 1979
- HOTELLING, H. u. M.R. PABST: Rank Correlation and Tests of Significance Involving no Assumption of Normality, in: Annals of Mathematical Statistics, 7/1936, S.29-43
- HULL, C.H. u. N.H. NIE: SPSS Update 7-9, New York 1981
- JOHNSON, N.L. u. S. KOTZ: Continuous Multivariate Distributions, 2 Bde., New York 1970
- KENDALL, M.G. u. A. STUART: The Advanced Theory of Statistics, London 1973
- KENNEDY, W.J. u. J.E. GENTLE: Statistical Computing, New York 1980

- KLEITER, F.: K-MICRO-STAT-1 (Programmbeschreibung),
Kiel 1981
- KOLBE, W.: Sortieren in Sekundenschnelle, in: CHIP 1/1982,
S.9
- KRICKEBERG, K.: Wahrscheinlichkeitstheorie, Stuttgart
1963
- KRUSKAL, W.H. u. W.A. WALLIS: Use of Ranks on One-Crite-
rion Variance Analysis, in: Journal of the Ameri-
can Statistical Association, 47/1952, S.583-621
- LEHMANN, E.L.: Testing Statistical Hypothesis, New
York 1959
- LEHMANN, E.L.: Nonparametrics. Statistical Methods Based
on Ranks, San Francisco 1975
- LING, R.F.: A Study of the Accuracy of Some Approximations
for the t , χ^2 , and F Tail Probabilities, in:
Journal of the American Statistical Association,
73/1978, S.274-283
- MAIORANA, J.A.: Changing a BASIC FOR...NEXT Loop into
a REPEAT...UNTIL Loop, in: BYTE 9/1981, S.162
- MANN, H.B.: Nonparametric Tests Against Trend, in:
Econometrika, 13/1945, S.245-259
- MANN, H.B. u. D.R. WHITNEY: On a Test of Whether One of
Two Random Variables is Stochastically Larger
Than the Other, in: Annals of Mathematical Statis-
tics 18/1947, S.50-60
- MAURER, A.: Datenstrukturen und Programmierverfahren,
Stuttgart 1974
- MEHLHORN, K.: Effiziente Algorithmen, Stuttgart 1977
- MOOD, A.M.: The Distribution Theory of Runs, in: Annals
of Mathematical Statistics, 11/1940, S.367-392
- MOORE, R.W.: Introduction to the Use of Computer
Packages for Statistical Analysis, Englewood
Cliffs 1978
- NELDER, J.A. u. R.C. MILTON: Statistical Computation,
New York 1969

- NEYMANN, J. u. E.S. PEARSON: On the Problem of the Most Efficient Tests of Statistical Hypotheses, in: Phil. Trans. Royal Soc. London 231/1933, S.289-337
- NIE, N.H. u. C.H. HULL: SPSS, New York 1975
- NORUSIS, M.J.: SPSS - Statistical Algorithms, Chicago 1979
- OSBORNE, A.: Einführung in die Mikrocomputertechnik, 3. Aufl., München 1978
- PEARSON, K.: On the Criterion that a Given System of Deviations from the Probable in the Case of Correlated System of Variables is such that it Can Be Reasonably Supposed to Have Arisen from Random Sampling, in: Philosophical Magazine Series 5, 50/1900, S.157-172
- PFANZAGL, J.: Allgemeine Methodenlehre der Statistik, o.O., o.J.
- RALSTON, A. u. S.H. WILF: Mathematische Methoden der Digitalrechner, Bd. 1, München/Wien 1970
- RICHTER, H.: Wahrscheinlichkeitstheorie, 2.Aufl., Berlin/Heidelberg/New York 1966
- ROSE, R.: Micro-SPSS. Statistisches Microcomputerprogramm für Sozialwissenschaften (Programmbeschreibung), Mohrkirch 1982
- SACHS, L.: Angewandte Statistik, Berlin/Heidelberg/New York 1974
- SAVAGE, I.R.: Bibliography of Nonparametric Statistics, Harvard University Press 1962
- SCHMETTERER, L.: Einführung in die Mathematische Statistik, 2. Aufl., Wien 1966
- SCHNELL, R.: Datenanalyse mit Kleinrechnern. Entwicklung eines interaktiven Datenanalysesystems für Micro-computer, Diplomarbeit Bochum 1982
- SIEGEL, S.: Nicht-parametrische statistische Methoden, Frankfurt a.M. 1976

- SMIRNOW, N.V.: Table for Estimating the Goodness of Fit of Empirical Distributions, in: Annals of Mathematical Statistics 19/1948, S.279-281
- TASSEL, D.V.: BASIC-Pack Statistics. Programs for Small Computers, Englewood Cliffs 1981
- TIEDE, M. u. W. VOSS: Induktive Statistik, 2 Bde., Köln 1977
- VELLEMAN, P.F. u. D.C. HOAGLIN: Applications, Basics, and Computing of Exploratory Data Analysis, Boston 1981
- WAERDEN, B.L. van der: Mathematische Statistik, Berlin 1971
- WALD, A. u. J. WOLFOWITZ: On a Test Whether Two Samples are from the Same Population, in: Annals of Mathematical Statistics, 11/1940, S.147-162
- WALLIS, W.A. u. G.H. MOORE: A Significance Test for Time Series Analysis, in: Journal of the American Statistical Association, 36/1941, S.401-409
- WILCOXON, F.: Individual Comparisons by Ranking Methods, in: Biometrics 1/1945, S.80-83
- WITTE, E.H.: Signifikanztest und statistische Inferenz, Stuttgart 1980
- WITTING, H.: Mathematische Statistik, Stuttgart 1966
- WITTING, H. u. G. NÖLLE: Angewandte Mathematische Statistik, Stuttgart 1970
- WOLFOWITZ, J.: On the Theory of Runs With Some Application to Quality Control, in: Annals of Mathematical Statistics, 14/1943, S.280-288

Anhang

I. Speicherplätze und Unterprogramm des APPLE II

Einige Programmteile benötigen spezielle Speicherplätze und Maschinenunterprogramme des APPLE. Für andere Rechner müssen eventuell Änderungen vorgenommen werden.

Unterprogramme

CALL-198	Es ertönt ein "Piep-Ton" (1 kHz) für 1/10 Sekunde.
CALL-958	Der Bildschirm wird unterhalb der aktuellen Cursorposition gelöscht.
CALL-998	Der Cursor wird an den Anfang der vorhergehenden Zeile gesetzt.

Speicherplatz Inhalt

32	linker Bildschirmrand; möglich sind 0 bis 39.
33	Bildschirmbreite in Zeichen. Wenn die Summe der Werte aus Speicher- stelle 32 und Speicherstelle 33 größer als 40 ist, können Störungen im Interpreter auftreten.
34	Oberer Bildschirmrand
35	Unterer Bildschirmrand. Der Wert in Speicherstelle 34 muß immer kleiner sein als der in Speicherstelle 35.

<u>Speicherplatz</u>	<u>Inhalt</u>
37	Zeile der aktuellen Cursorposition
216	Errorflag. Wenn PEEK (216) < 218 ist, so wurde kein Errorflag gesetzt.
218-219	Die Zeilennummer, an der ein Fehler aufgetreten ist, läßt sich feststellen durch: PEEK (218) + PEEK (219)·256
222	Fehlercode eines aufgetretenen Fehlers
-16384	ASCII-Code des zuletzt eingetippten Zeichens. Ist PEEK (-16384) > 128, so wurde eine Taste gedrückt.
-16368	Damit ein neues Zeichen eingelesen werden kann, muß diese Speicherstelle auf Null gesetzt werden.

II. Inhaltsverzeichnis der Systemdiskette

Catalog des REP-TEST

VOLUME=254		FREE SECTORS=220 (55K)	
PT	LEN	START	BLEN NAME
A	2		194 HELLO
*A	10		2141 REP-TEST MENU
*A	20		4955 FILE INPUT
*A	20		6899 DATA INPUT
*A	60		15090 DATA BASE SYSTEM
*A	79		19947 TESTVERFAHREN
*A	76		19054 TESTS MIT HILFEN

Erläuterung:

P = Protected: Die Programme sind gegen versehentliches Löschen geschützt

T = Type A bedeutet APPLESOFT-BASIC-Programm
B bedeutet binäres File

LEN = Length: Länge des Files in Sektoren

START = Startadresse eines Maschinenprogramms

BLEN = Binary Length:Länge des Files in Bytes

III. Programm Listings

REP-TEST MENU	231
TESTVERFAHREN	233
TESTS MIT HILFEN	250
FILE INPUT	265
DATA INPUT	269
DATA BASE SYSTEM	274


```

*****
*                                     *
*               ***REP-TEST***       *
*                                     *
*   NICHT-PARAMETRISCHE TESTVERFAHREN *
*                                     *
*           (C) RENATE PRUST 1993     *
*                                     *
*               REP-TEST MENU        *
*                                     *
*****

```

```

10 DATA FILE INPUT,7,DATA INPUT,19,DATA BASE SYSTEM,17,TESTVERFAHREN,
EN,21,TESTS MIT HILFEN,20: DIM P$(5),L$(5): ONERR GOTO 60000
20 D$ = CHR$(4): FOR I = 1 TO 5: READ P$(I),D$(I): NEXT I
GOTO 50
30 HOME: VTAB 10: HTAB 9: PRINT "LOADING...": P$=A$: HTAB
14: HTAB 10: PRINT "DAUER: ";D$(A%): " SEKUNDEN": RETURN
50 POKE 34,0: HOME: HTAB 13: PRINT "****REP-TEST**": PRINT
: HTAB 4: PRINT "NICHT-PARAMETRISCHE TESTVERFAHREN": PRINT:
PRINT: POKE 34,2
60 HTAB 4: PRINT "FILE INPUT.....1": PRINT:
: HTAB 4: PRINT "DATA INPUT.....2": PRINT:
HTAB 4: PRINT "DATA BASE MANAGEMENT SYSTEM.....3": PRINT:
70 HTAB 4: PRINT "TESTVERFAHREN.....4": PRINT:
: HTAB 4: PRINT "TESTS MIT ENTSCHEIDUNGSHILFEN...5": PRINT:
HTAB 4: PRINT "ENDE DES PROGRAMMS.....6": PRINT
80 HTAB 4: INPUT "GETROFFENE AUSWAHL.....": A$: IF
A% < 1 OR A% > 6 THEN CALL - 998: CALL - 958: GOTO 80
90 IF A% = 6 THEN HOME: END
100 ON A% - 1 GOTO 120,130,140,150
110 GOSUB 30: PRINT D$"RUN FILE INPUT"
120 GOSUB 30: PRINT D$"RUN DATA INPUT"
130 GOSUB 30: PRINT D$"RUN DATA BASE SYSTEM"

```

```

140 HOME : HTAB 6: PRINT "DIREKTE AUSWAHL DER VERFAHREN": PRINT
: PRINT : PRINT "SIE KOENNEN GLEICH DEN GEWUENSCHTEN": PRINT
"TEST AUSWAEHLEN.": PRINT "FALLS SIE NOCH NICHT WISSEN, WELCHES":
PRINT "VERFAHREN SIE ANWENDEN KOENNEN,"
150 PRINT "GEBEN SIE EIN ? EIN UND WAEHLEN SIE": PRINT "AUS
DEM MENU NUMMER 5.": PRINT : PRINT "SONST GEBEN SIE ! EIN."
160 VTAB 15: HTAB 27: PRINT "EINGABE: ";: GET A$: IF A$ :
> "!" AND A$ < > "?" THEN CALL - 998: CALL - 958: GOTO 160
170 PRINT A$: IF A$ = "?" THEN 50
180 GOSUB 30: PRINT D$"RUN TESTVERFAHREN"
190 HOME : HTAB 11: PRINT "AUSWAHL MIT HILFEN": PRINT : PRINT
: PRINT "SIE WERDEN UEBER MEHRERE SCHRITTE ": PRINT "ZU EINEM
VERFAHREN GEFUEHRT, DASS DER": PRINT "RECHNER NACH IHREN ANGABEN
AUSWAHLT."
200 PRINT "FALLS SIE SCHON WISSEN, WELCHES VERFAH-": PRINT
"REN EINGESETZT WERDEN KANN, GEBEN SIE": PRINT "EIN ? EIN UND
WAEHLEN SIE AUS DEM": PRINT "DIE NUMMER 4."
210 PRINT "SONST GEBEN SIE EIN ' EIN."
220 VTAB 15: HTAB 27: PRINT "EINGABE: ";: GET A$: IF A$ :
> "!" AND A$ < > "?" THEN CALL - 998: CALL - 958: GOTO 220
230 PRINT A$: IF A$ = "?" THEN 50
240 GOSUB 30: PRINT D$"RUN TESTS MIT HILFEN"
299 REM

```

```

60000 IF PEEK (222) = 254 THEN VTAB 22: CALL - 198: INVERSE
: HTAB 4: PRINT "UNZULAESSIGE EINGABEWIEDERHOLUNG": FOR Q =
1 TO 1000: NEXT : NORMAL : VTAB 18: CALL - 958: GOTO 80
60010 HOME : IF PEEK (222) = 6 THEN VTAB 12: CALL - 198:
HTAB 10: PRINT "FALSCHE DISKETTE": CLEAR : FOR Q = 1 TO 1000:
NEXT : GOTO 10
60020 IF PEEK (222) = 8 THEN VTAB 12: CALL - 198: HTAB 6:
PRINT "LESEN/SCHREIBEN UNMOEGLICH": CLEAR : FOR Q = 1 TO 1000:
NEXT : GOTO 10
60030 VTAB 9: CALL - 958: PRINT CHR$ (7): VTAB 10: HTAB 12:
PRINT "FATAL ERROR": VTAB 12: HTAB 10: PRINT "NEUSTART DES SYSTEMS":
VTAB 22: END

```

Programme I : TESTVERFAHREN


```

*****
*                                     *
*               ***REP-TEST***       *
*                                     *
*   NICHT-PARAMETRISCHE TESTVERFAHREN *
*                                     *
*           (C) RENATE PRUST 1983     *
*                                     *
*           TESTVERFAHREN             *
*                                     *
*****

```

```

10  DIM IO(201),A(5,201),X(201),Y(201): DIM V$(82),VC$(5),VM$(5):
DIM TX(9),S(9),NX(9,9): DEF FN STN(Z) = EXP (- (.83 * ABS
(Z) + 351) * ABS (Z) + 562) * ABS (Z) / (.03 + 1.5 * ABS
(Z))) / 2:PI = 3.141592653: ONERF GOTO 61000
20  DEF FN R(Z) = INT (Z * 1000 + .5) / 1000:IC% = 999:F2$
= " VARDATA FILE ": HOME : CALL - 198: VTAB 10: HTAB 5: INVERSE
: PRINT "LEGEN SIE DIE DATENDISKETTE EIN": PRINT : HTAB 3: PRINT
" UND DRUECKEN SIE DIE RETURN-TASTE": NORMAL : GET A$: PRINT
: HOME
30  VTAB 10: HTAB 10:A$ = "VARCATALOG": PRINT "READING...":A$:
PRINT : HTAB 8: PRINT "DAUER: MAX. 15 SEKUNDEN": PRINT CHR$
(4)"OPEN":A$: PRINT CHR$(4)"READ":A$: INPUT F$,NU%,NF%,MW,IS%,IN$:
FOR J = 1 TO NV%: INPUT V$(J): NEXT : PRINT CHR$(4)"CLOSE":A$:
GOTO 1000
100 H1 = N - 1
110 H1 = INT ((H1 - 2) / 3) + 1: FOR H2 = 1 TO N - H1:H3 =
H2 + H1:H4 = X(H3):H5 = Y(H3): IF X(H2) < = H4 THEN 160
120 H6 = H2
130 X(H3) = X(H6):Y(H3) = Y(H6):H3 = H6: IF H6 < = H1 THEN
150
140 H6 = H6 - H1: IF X(H6) > H4 THEN 130
150 X(H3) = H4:Y(H3) = H5
160 NEXT H2: IF H1 > 1 THEN 110
170 RETURN

```

```

100 H4% = 0:H2% = 0:H3% = 0: IF X(1) = 1:CN: THEN FOR J = 1 TO
110 N: H4% = H4% + 1: 2: NE 1: H4% = H4% + N:H3% = H4% * N - N:H2% =
120 H4% - N:H4% = N: RETURN
130 CN = 1: X(1) = X(1):S = 1: FOR J = 1 TO N: IF H4% = H4% + 1
140 THEN NE 1
150 H1 = H3 + J: / 2: FOR Q = 5 TO J: D = H1: H5 = H4 +
160 H1 + H4: H4 = H4 + H4:H2 = H3 + H1 + H4: 4:H4 = H1 + H4:H4%
170 = H4% + H4 + 1: 5: H4 = H4 + 1: S = J + 1: NEXT J: H3 = H4
180 S = 0: IF V(1) < 34 THEN S = FN R(1) + FN R(1) + FN R(1) + FN R(1)
190 S = SQR(X(1) + T + T / 2) + H1: RETURN
200 T = ABS(T) + H3 = 1:H3 = T: SQR = H1 + ATN(H3):H4 =
210 H3 + SQR(1 + H3 + H3 + H3 + H3 = H4: H3: IF H3 = 1 THEN S
220 = FN R(1) - 2 * H3: GOTO 1: RETURN
230 H1% = 0 - 3:H2% = INT(X(1) / 2) + 1: H3 = 2 - H2%: IF
240 H3 < H1% THEN FOR H1 = H3 TO H1% STEP 1: S = H3 + H1 + H1
250 + 1: S = S + H3 * H3: H1 = H1 + 1: NEXT
260 IF NOT H2% THEN S = FN R(1 - ((H5 + S) * H6 + H7) * 2
270 / PI): RETURN
280 S = FN R(1 - H6 * S - H6): RETURN
290 IF CH > 90 OR CH < .0000001 OR V > 30 THEN S = 100: RETURN
300 H1 = 1: FOR H2 = 1 TO 1: STEP .2:H1 = H1 + H2: NEXT H2: H3
310 = CH / INT(CH + 1) / 2 + 1: SQR = CH / 2 + H1:H4 = 1: 1:
320 INT(V / 2) + 1: IF H1 > SQR(2) + H1: PI
330 H7 = .00000001:H5 = 1:H6 = 1: FOR Q = 0 TO 1: H1 = 1 + 2:H6
340 = H6 * CH / 10:H5 = H5 + H6:H4 = H4 + H7: NEXT Q: IF H1 =
350 H4 * H3 * H5: RETURN
360 Q = 1 - F:H6 = 1:H7 = 0: 4: H1 = FN R(H5 + H7: IF H1 >
370 THEN RETURN
380 FOR J = 1 TO K:H2 = J: GOSUB 310:H6 = H6 * F:H7 = H7
390 S = S + H1 * H6 * H7: NEXT J: S = FN R(S): RETURN
400 H1 = 1: IF H2 = 0 OR H2 = H2 THEN RETURN
410 H2 = 1: IF H2 = H2: H3 = H2: H4 = H3 + 1
420 FOR H3 = 1 TO H3 - H2:H1 = H1 + H3: H2 = H3: FOR H3 = H3
430 + 1 TO H3:H2 = H2 * H3: NEXT H3: H1 = H1 + H2: H3 = H3 + 1
440 VTAB 15: HTAB 10: PRINT "LOAD: C...": GOTO 1: IF H1 > 100
450 PRINT "DAUER: 7 SEKUNDEN"
460 A$ = V$(V$(I%)) + F2$ + F$: PRINT CHR$(147) + BLOWO(A$): GOTO
470 PEEK(107) + PEEK(108) * 256: FOR Q = 1 TO 100: H1 = 1: 1:
480 NEXT: RETURN
490 H1 = 5: FOR J = 1 TO 100: VTAB (H1 + J): PRINT "VARIABLE
500 ";J: INPUT "":A$: GOSUB 440:V$(J) = A$: GOTO 490 + 1:
510 HTAB 20: PRINT V$(V$(J)): NEXT J: FOR J = 1 TO 100: GOSUB 470:
520 NEXT J: FOR J = 1 TO 100: FOR J = 1 TO 100: IF V$(J) = V$(J)
530 THEN A$(J) = J
540 NEXT: NEXT: RETURN
550 HTAB 6: INPUT "GEWUNSCHE VARIABLE ";A$
560 IF A$ = "" THEN FOR J = 1 TO 100: GOTO 1000
570 I% = 0: FOR I = 1 TO 100: IF A$ < V$(I) THEN NEXT:
580 VTAB 20: HTAB 5: PRINT "VARIABLE ";A$: PRINT "EXISTIERT
590 NICHT": FOR Q = 1 TO 1500: NEXT: CALL - 998: CALL - 956:
600 GOTO 430
610 I% = I: RETURN

```

```

420 FOR I = 1 TO NCX: IF VMX(J) = VMX(I) THEN RETURN
480 NEXT : IF NCX < 5 THEN NCX = NCX + 1:IX = NCX: GOTO 410
450 FOR I = 1 TO NCX: FOR K = 1 TO VCX: IF VMX(K) = VMX(I)
THEN NEXT I
500 NEXT :IX = 1
510 VCX, IX = VMX(J): GOSUB 400: RETURN
520 CX = 0: FOR J = 1 TO NFV: X = A(VMX(1),J): IF X < > MW THEN
IF Y < > MW AND X < > MW THEN CX = CX + 1: X CX = X: Y = Y
530 NEXT : RETURN
540 CX = 0: FOR J = 1 TO NFV: X = A(VMX(1),J): IF X < > MW THEN
CX = CX + 1: X CX = X
550 NEXT : RETURN
560 IX = 0: FOR J = 1 TO NFV: FOR I = 1 TO VCX: X = A(VMX(I),J):
IF X < > MW THEN NEXT : CX = CX + 1: FOR I = 1 TO VCX: A(VMX(I),J)
= A(VMX(I),J): NEXT
570 NEXT J: RETURN
580 HOME : H1% = LEN (A$:H2% = INT ((40 - H1%) / 2): FOR
J = 1 TO H2%: PRINT "-": NEXT : PRINT A$:: FOR J = H2% + 1
+ 1 TO 40: PRINT "-": NEXT : POKE 34,4: HOME : RETURN
610 HOME : UTAB 10: HTAB 14: PRINT "RECHNUNG": PRINT : HTAB
3: PRINT "DAUER: MAX. ";DUX + 1:" SEKUNDEN": RETURN
620 HOME : UTAB 5: PRINT "VARIABLE ";V$(VCX(VMX(1))): RETURN

630 HOME : PRINT "VARIABLE ";V$(VCX(VMX(1))):" MIT VARIABLE
";V$(VCX(VMX(2))): RETURN
640 HOME : PRINT "TESTVARIABLEN": PRINT : FOR J = 1 TO VCX:
PRINT V$(VCX(VMX(J))): NEXT : RETURN
650 PRINT : PRINT "GUELTIGE FAELE ";CX:"" FAELENDE FAELE
":NFV - CX: RETURN
660 FOR J = 0 TO 9:S(J) = 0:TX(J) = 0: FOR I = 0 TO 9:N%(I,J)
= 0: NEXT : NEXT : RETURN
670 H1 = 0:H2 = 0:H3 = 0:H4 = 0:H5 = 0:H6 = 0:H7 = 0:H8 = 0:H9
= 0:H1% = 0:H2% = 0:H3% = 0:H4% = 0:H6% = 0:H7% = 0:H8% = 0:H9%
= 0: RETURN
700 UTAB 24: HTAB 6: INVERSE : PRINT "DRUECKEN SIE IRGENDNE
TASTE": NORMAL : POKE - 10368,0: WAIT - 16384,128: POKE
- 10368,0: PRINT : RETURN
710 HOME : UTAB 10: CALL - 193: PRINT "KEINE GUELTIGEN FAELE.
ABBRUCH": GOSUB 700: GOTO 1000
720 FOR J = 1 TO CX:H1 = Y(J): IF H1 = 0 OR H1 = 1 THEN NEXT
: RETURN
730 HOME : UTAB 10: CALL - 193: PRINT "FEHLER: VARIABLE NICHT
0,1 CODIERT.": POP : GOSUB 700: GOTO 1000
740 HOME : UTAB 8: CALL - 193: PRINT "STANDARDABWEICHUNG =
0. ABRUCH ": GOSUB 700: GOTO 1000
750 PRINT "CODE ANZAHL MITTL. RANG": PRINT : FOR I = 0 TO
H1X: IF TX(I) < > 0 THEN HTAB 2: PRINT I:: HTAB 9: PRINT TX(I)::
HTAB 16: PRINT FN R(S(I) / TX(I))
760 NEXT : PRINT : RETURN

```

```

770 S = 0: IF NX(0,1) * NX(1,0) > NX(0,0) * NX(1,1) THEN X =
NX(0,0):NX(0,0) = NX(1,0):NX(1,0) = NX(0,1):NX(0,1) =
NX(1,1)
780 H9 = NX(0,0) + NX(1,0):H8 = NX(0,0): GOSUB 310:D = H1:H5
= CX - H7:H8 = NX(0,1): GOSUB 310:Q = 0 * H1:H9 = CX:H8 = NX(0,0)
+ NX(0,1): GOSUB 310:S = S + Q / H1: IF (H1 - 0.1) = 0 AND (H1 - 0)
> 0 THEN NX(0,0) = NX(0,0) + 1:NX(0,1) = NX(0,0) - 1:NX(1,0)
= NX(1,0) - 1:NX(1,1) = NX(1,1) + 1: GOTO 780
790 PRINT "FISHER'S WAHRSC.": HTAB 25: PRINT FN R(S): PRINT
: RETURN
800 N = CX: GOSUB 100:H6 = 0: 2:H9% = H5 + .5:H9 = X/H9%:
IF INT (H6) = H6 THEN H9 = (X/H6) + X/H6 + 1: 2
810 FOR J = 1 TO CX:X = 0: IF X(D) > H9 THEN X = 1
920 X(J) = X: NEXT : RETURN
830 FOR J = 1 TO CX:Y = Y(D) - H1:TX = X(J):H2 = Y(J)
+ 1:TX(X) = TX(X) + 1:S(Y) = S(Y) + 1: NEXT : GOSUB 630: PRINT
: RETURN
840 FOR I = 0 TO H3%: HTAB 15: PRINT "CODE ":I + H1%: HTAB
26: PRINT NX(0,I): HTAB 31: PRINT NX(1,I): HTAB 36: PRINT
S(I): NEXT : HTAB 26: PRINT TX(0): HTAB 31: PRINT TX(1): HTAB
36: PRINT CX: PRINT : RETURN
850 PRINT "MEDIAN: ": FN R(H9): HTAB 25: PRINT " (P = ) ":
RETURN
860 H3 = Y(1) - X(1):H4 = H3: FOR J = 2 TO N:H2 = Y(J) - X(J):
IF H2 > H3 THEN H3 = H2
870 IF H2 < H4 THEN H4 = H2
880 IF A% = 3 THEN 910
890 H1 = Y(J - 1) - X(J): IF H1 > H3 THEN H3 = H1
900 IF H1 < H4 THEN H4 = H1
910 NEXT :H5 = ABS (H3): IF ABS (H4) > H5 THEN H5 = ABS
(H4)
920 S = 0:Z = H5 * SQR (N): IF Z < .27 THEN S = 1
930 IF Z > = .27 AND Z < 1 THEN Q = EXP ( - 1.233701 / Z
/ Z):H7 = Q ^ 9:S = 1 - 2.506628 / Z * (Q + H7 + H7 * H7 * H7
/ Q / Q)
940 IF Z > = 1 AND Z < 3.1 THEN Q = EXP ( - 2 * Z * Z):H7
= Q ^ 4:S = Z * (Q - H7 + H7 * H7 / Q - H7 * H7 * H7 * H7)
950 RETURN
960 HTAB 6: PRINT "ABS. DIFF. POS. DIFF. NEG. DIFF.": PRINT
: PRINT "MAX.": HTAB 8: PRINT FN R(H5): HTAB 20:A$ = "-":
IF H3 > = 0 THEN A$ = STR$ ( FN R(H3))
970 PRINT A$: HTAB 32:A$ = "-": IF H4 < 0 THEN A$ = STR$
( FN R(H4))
980 PRINT A$: PRINT : PRINT : PRINT "TESTWERT Z": HTAB 25:
PRINT FN R(Z): PRINT : PRINT "SIGNIFIKANZ": HTAB 25: PRINT
FN R(S): RETURN

```

```

1000 YY = FRE (0): POKE 34,2: HOME : PRINT "BINOMIAL.....1
  CHI QU.-UNABH....11": PRINT : PRINT "CHI QU.-1 ST....2  SPEARMAN...
  ....12": PRINT : PRINT "KOLMOG./SM.-1 ST.3  BRAVAIS/PEARSON...13"
1020 FOR J = 1 TO 39: PRINT "-";: NEXT : PRINT : PRINT "CHI
QU.-MEHR ST..4  MEDIAN.....14": PRINT : PRINT "FISHER.....
5  MANN/WHITNEY....15": PRINT : PRINT "KOLM./SM.-2 ST...6
KRUSKAL/WALLIS...16": PRINT
1030 PRINT "WALD/WOLFWITZ...7  DIXON/MOOD.....17": PRINT
: PRINT "MCNEMAR.....8  WILCOXON.....18": PRINT : PRINT
"COCHRAN.....9  FRIEDMAN.....19"
1040 FOR J = 1 TO 39: PRINT "-";: NEXT : PRINT : PRINT "RUNS..
...10  REP-TEST MENU....20": PRINT : POKE 34,22: HOME
1050 INPUT "AUSWAHL.....":AX: IF VAL ( STR$ (AX)) <
A: OR AX < 1 OR AX > 20 THEN HOME : GOTO 1050
1060 POKE 34,2: IF AX < > 20 THEN ON AX GOTO 11000,12000,13000,1400
0,15000,16000,17000,18000,19000,20000,21000,22000,23000,24000,25000,26
000,27000,28000,29000
1070 HOME : VTAB 10: HTAB 3: INVERSE : PRINT "LEGEN SIE DIE
SYSTEMDISKETTE EIN": PRINT : HTAB 3: PRINT "UND DRUECKEN SIE
DIE RETURN-TASTE";: NORMAL : INPUT "":A$
1080 HOME : VTAB 12: HTAB 12: PRINT "ZURUECK ZUM MENU": PRINT
CHR$ (4)"RUN REP-TEST MENU"
1100 REM

```

```

11000 A$ = "BINOMIAL": GOSUB 600:VC% = 1: GOSUB 420: GOSUB 540:
HOME : IF C% = 0 THEN 710
11010 FOR J = 1 TO C%:Y(J) = X(J): NEXT : GOSUB 720: VTAB 7:
PRINT "WAHRSCH. CODE <1/>": HTAB 26: INVERSE : PRINT "VOREINST.=0.5";:
NORMAL
11020 HTAB 18: INPUT A$:P = .5: IF A$ < > "" THEN P = VAL
(A$): IF P < 0 OR P > 1 THEN CALL - 998: CALL - 958: GOTO
11020
11030 DUX = NF% / 16: GOSUB 610:H9% = 0: FOR J = 1 TO C%:H9%
= H9% + X(J): NEXT : IF C% < 26 THEN H9 = C%:K = H9%: GOSUB
290: GOTO 11050
11040 Z = (H9% - SGN (H9% - C% * P) * .5 - C% * P) / SQR (C%
* P * (1 - P)):S = FN R( FN STN(Z))
11050 GOSUB 620: VTAB 8: PRINT "CODE  HAEUF.  REL.  WAHRSCH.":
PRINT : PRINT " 0": HTAB 10: PRINT C% - H9%: HTAB 18: PRINT
FN R(1 - H9% / C%): HTAB 26: PRINT FN R(1 - S): PRINT : PRINT
" 1": HTAB 10: PRINT H9%: HTAB 18: PRINT FN R(H9% / C%):
HTAB 26: PRINT S: PRINT : PRINT
11060 PRINT "BERNOULLI-PARAMETER ":P: PRINT : GOSUB 650: GOSUB
700: GOTO 1000

```

```

12000 A$ = "CHI QU.-ANPASSUNGSETEST": GOSUB 400: CH = 1: GOSUB
420: GOSUB 540: IF C% = 0 THEN 710
12010 GOSUB 670: DUX = C% / 10 + 1: GOSUB 500: H1% = H2%
+ H3%: FOR I = 2 TO CH: X = X * J: IF H1% = 0 THEN H1% =
1000: IF H3% = 1 THEN H2% = X
12030 NEXT I: H3% = H2% - H1% + 1: IF H3% > 1400 THEN 10
PRINT "TUMULE KATEGORIEN, ABRUCH": GOSUB 210: GOTO 1000
12040 HOME: INPUT "PRUEFUNG AUF GLEICHHEITEN: JA? > 1: N? > 2:": A$:
IF A$ < > "1" AND A$ < > "2" THEN CALL = 99: CALL = 99:
GOTO 12040
12050 FOR J = 1 TO CH: X = X * J: H1% = H1% + 1: H2% = H1% + 1: NEXT
: IF A$ = "2" THEN FOR I = 0 TO H3% - 1: H1% = H1% + 1: H2% = H2% + 1:
: GOTO 12070
12060 HOME: PRINT "CODE BEOBSCHTET": PRINT: PRINT: PRINT:
Y = 0: FOR I = 0 TO H3% - 1: HTAB 1: PRINT I + H1%: HTAB 11:
PRINT TX(I): HTAB 23: INPUT S(I): I = I + 1: IF I = 14
+ 1: CH THEN CALL = 128: PRINT: PRINT "SUMME DER ERWARTETEN
HAUEFIGK.": PRINT "ANZAHL DER GUELTIGEN: HELL": GOSUB
700: GOTO 12060
12070 GOSUB 600: CH = 0: FOR I = 0 TO H3% - 1: CH = CH + TX(I)
+ S(I): I = I + 1: IF S(I) = 5 THEN HOME: PRINT "ABBRUCH":
GOSUB 100: GOTO 1000
12080 NEXT I: H2% = H2% - 1: GOSUB 200: GOSUB 400: PRINT: PRINT
"CODE BEOBSCHTET ERWARTET": PRINT: FOR I = 0 TO H3% -
1: HTAB 2: PRINT I + H1%: HTAB 12: PRINT "TX(I)": HTAB 13: PRINT
FN R(S(I)): NEXT I: PRINT
12090 PRINT "FREIHEITSGRADE": HTAB 25: PRINT "CH" = CH: PRINT
"TESTWERT CH": HTAB 25: PRINT FN R(CH): PRINT "SIGNIFIKANZ":
HTAB 25: PRINT S: GOSUB 650: GOSUB 700: GOTO 1000

```

```

13000 A$ = "KOLMOGOROFF/SMIRNOW-1 STICHPROBEN-TEST": GOSUB 400:UCK
= 1: GOSUB 420: GOSUB 540: HOME: IF CX = 0 THEN 710
13010 GOSUB 670: DUX = CX / 5: PRINT "WELCHE VERTEILUNG WIRD
ALS": PRINT: PRINT "THEORETISCHE VERTEILUNG ANGENOMMEN": POKE
32,4: POKE 33,35: PRINT: PRINT: PRINT "GLEICHVERTEILUNG.....1":
PRINT: PRINT "NORMALVERTEILUNG.....2": PRINT: PRINT "POISSONVERTEILU
NG.....3": PRINT
13020 INPUT "AUSWAHL.....": H1: IF A$ < 1 OR A$ > 3
THEN CALL - 998: CALL - 958: GOTO 13020
13030 POKE 32,0: POKE 33,40: GOSUB 610: N = CX: GOSUB 1000: N
= 0: Y = X(1): FOR J = 1 TO CX: X = X(J): H9 = H9 + X: H8 = H8 +
X * X: IF X < > Y THEN N = N + 1: Y(N) = (J - 1) / CX: =
13040 NEXT: H9 = H9 / CX: N = N + 1: Y(N) = 1: ON (N) GOTO (13000,13010,
13020): GOSUB 670: Y(0) = 0: V = CX: GOSUB 360: GOSUB 620: PRINT
: PRINT: GOSUB 950: GOSUB 650: VTAB 5: HTAB 20: PRINT V$(0):
GOSUB 700: GOTO 1000
13100 DUX = CX / 10: HOME: VTAB 6: PRINT "WOLLEN SIE MINIMUM
UND": PRINT: PRINT "MAXIMUM ANGEBEN (J/N) ? ": INPUT "": A$:
IF A$ = "N" THEN H1 = X(1): H2 = X(CX): GOTO 13120
13110 PRINT: PRINT: INPUT "MINIMUM ? ": H1: PRINT: PRINT
"MAXIMUM ? ": H2
13120 GOSUB 610: Q = 0: Y = X(1) - 1: FOR J = 1 TO CX: X = X(J):
IF X < > Y THEN Q = Q + 1: X(Q) = X - H1: H2 = H2 + H1: =
13130 NEXT: V$(0) = "GLEICHVERT.": RETURN
13200 DUX = CX / 10: HOME: VTAB 8: PRINT "WOLLEN SIE PARAMETER
ANGEBEN ?": PRINT: INPUT "MITTELWERT ? ": A$: IF A$
= "" THEN H9 = VAL (A$)
13210 PRINT: INPUT "STANDARDABWEICHUNG ? ": A$: IF A$ < 1
OR A$ > 10 THEN H7 = VAL (A$): GOTO 13230
13220 H7 = SQR ((H8 - H9 * H9 * CX) / (CX - 1)): IF H7 = 0
THEN POP: GOTO 740
13230 GOSUB 610: Q = 0: Y = X(1) - 1: FOR J = 1 TO CX: X = X(J):
IF X < > Y THEN Q = Q + 1: Z = (X - H9) / H7: X(Q) = FN STN(2, 0,
= 1: IF Z > 0 THEN X(Q) = 1 - X(Q)
13240 NEXT: V$(0) = "NORMALVERT.": RETURN
13300 FOR J = 1 TO CX: X = X(J): IF X < > INT (X) OR X < 0
THEN HOME: VTAB 9: PRINT "DIE DATEN MUESSEN GANZZAHLEIGENDIG":
PRINT "NICHT-NEGATIV SEIN. ABRUCH": POP: GOSUB 700: GOTO 1000
13310 DUX = CX / 10: HOME: VTAB 8: PRINT "WOLLEN SIE DEN MITTELWERT":
PRINT: INPUT "EINGEBEN ? (J/N) ": A$: IF A$ = "J" THEN PRINT
: PRINT: INPUT "MITTELWERT: ": H9
13320 GOSUB 610: IF H9 = 0 THEN HOME: VTAB 10: PRINT "MITTELWERT
= 0. ABRUCH": POP: GOSUB 700: GOTO 1000
13330 H1 = EXP (- H9): Q = 1: H3 = Q: H4 = H3 * X(Q): IF H1
< H4 THEN FOR I = 1 TO X: H3 = H3 * H9 / I: H4 = H4 + H3: NEXT
13340 X(Q) = H4 * H1: Y = X: FOR J = 2 TO CX: X = X(J): IF Y
< X THEN FOR I = Y + 1 TO X: H3 = H3 + H9 / I: H4 = H4 + H3:
NEXT: Q = Q + 1: X(Q) = H4 * H1: Y = X
13350 NEXT: V$(0) = "POISSONVERT.": RETURN

```

```

14300 A$ = "CHI QU.-HOMOGENITAETSTEST": GOSUB 300:V% = 2: PRINT
"HINWEIS: DIE VARIABLEN MUESSEN": PRINT "GANGBAR, VON 0 AUFWAERTS":
PRINT "SIS HOECHSTENS 9, CODIERT SEIN": GOSUB 700: HOME : GOSUB
420: GOSUB 520: HOME : IF C% = 0 THEN 710
14310 GOSUB 670: GOSUB 860:DUM = C% 4: GOSUB 110: FOR I =
1 TO C%:X = X(J):Y = Y(J): IF X > 9 OR Y > 9 THEN PRINT "FALSCHE
CODIERUNG. ABBRUCH": GOTO 1000
14320 NX(X,Y) = NX(X,Y) + 1:TX(X) = TX(X) + 1:SY(Y) = SY(Y) +
1: IF X > H1% THEN H1% = X
14330 IF Y > H2% THEN H2% = Y
14340 NEXT :CH = 0:V = H1% * H2%: IF V = 1 AND C% < = 30 THEN
GOSUB 770: GOTO 14080
14350 IF V = 1 AND C% > 30 THEN CH = C% / ( ABS (NX(0,0) +
NX(1,1) - NX(0,1) * NX(1,0)) - C% / 2) ^ 2 / TX(0) / SY(0) /
SY(1): GOTO 14070
14360 FOR I = 0 TO H1%: FOR J = 0 TO H2%:CH = CH + (NX(I,J)
- SY(J) * TX(I) / C% ^ 2 / SY(J) / TX(I) * C%: NEXT : NEXT
14370 GOSUB 240: GOSUB 630: PRINT : PRINT : PRINT "FREIHEITSGRADE":
HTAB 25: PRINT H1% * H2%: PRINT : PRINT "TESTWERT CHI": HTAB
25: PRINT FN RICH: PRINT : PRINT "SIGNIFIKANZ": HTAB 25:
PRINT S
14380 GOSUB 450: GOSUB 700: GOTO 1000
14390 REM
14399 REM

```

```

15000 A$ = "FISHER-TEST": GOSUB 600: PRINT "HINKEISCHVARIABLEN  

MUESSSEN 0/1 DOOBIERT SEIN": GOK = 2: GOSUB 700: HOME : GOSUB 400:  

GOSUB 520: GOK = GOK / 5: GOSUB 610: IF GOK = 0 THEN GOTO  

15010: GOSUB 720: FOR J = 1 TO C%*(40/J) = 400/J: J = J/J:  

NEXT : GOSUB 730: FOR J = 1 TO C%*(40/J) = 400/J: NEXT : GOSUB  

400: GOSUB 670: IF GOK < 31 THEN 15020  

15015 A$ = "CHI QUANT-HOMOGENITAETSTEST": CALL = 198: HOME :  

HTAB 9: HTAB 4: PRINT "N IST ZU GROSS, ES WIRD EIN": PRINT  

: HTAB 4: PRINT A$: PRINT : HTAB 4: PRINT "DURCHGESCHLEHT":  

FOR J = 1 TO 2000: NEXT : POKE 34,2: HOME : GOSUB 600: GOTO  

14010  

15020 H3% = 1: GOSUB 800: GOSUB 630: PRINT : HTAB 9: PRINT "VERWURFEN":  

: HTAB 25: PRINT "40 / 1": PRINT : PRINT "VERWURFEN": GOSUB  

840: GOSUB 770: GOSUB 650: GOSUB 700: GOTO 1000

```



```

16000 A# = "KOLMOGOROFF/SMIRNOW-2 STICHPROBEN-TEST": GOSUB 600:VC%
= 2: PRINT "HINWEIS:VARIABLE2 MUSS 0/1 CODIERT SEIN.": GOSUB
700: HOME : GOSUB 420: GOSUB 520: IF CX = 0 THEN 710
16010 GOSUB 720: GOSUB 670:DU% = CX / 4: GOSUB 610:N = CX:
GOSUB 100: GOSUB 660: FOR J = 1 TO CX:Y = Y(J):TX(Y) = TX(Y)
+ 1: NEXT :H8% = TX(0):H9% = TX(1):H1% = 0:N = H8%: IF H9% >
H8% THEN H1% = 1:N = H9%
16020 Q = CX - N:H2% = ABS (H1% - 1):H1 = Y(CX) + 1: FOR
= CX TO 1 STEP - 1:X = X(J):Y = Y(J): IF X = H1 THEN X(J) =
X(J + 1):Y(J) = Y(J + 1): GOTO 16040
16030 Y(J) = TX(H1%) / N:X(J) = TX(H2%) / Q
16040 TX(Y) = TX(Y) - 1:H1 = X: NEXT :AX = 3:N = CX:H = H2%
* H9% / CX: GOSUB 860: GOSUB 630: PRINT : PRINT "CODE 0: " :HTAB 18:
PRINT "CODE 1: " :H9%: PRINT : GOSUB 960: GOSUB 650:
GOSUB 700: GOTO 1000
16999 REM

```

```

17000 A# = "WALD/WOLFWITZ-RUNTEST": GOSUB 600: PRINT "HINWEIS:WALD-
E2 MUSS 0/1 CODIERT SEIN":VC% = 2: GOSUB 700: HOME : GOSUB 420:
GOSUB 520: HOME : IF CX = 0 THEN 710
17010 GOSUB 720:DU% = CX / 5:S = 0: GOSUB 670: GOSUB 610:N
= CX: GOSUB 100:H2% = Y(1):H4% = 0: FOR J = 2 TO CX:H2% = H2%
+ Y(J): IF Y(J) < > Y(J - 1) THEN H9% = H9% + 1: IF AX% =
X(J - 1) THEN H4% = 1
17020 NEXT :H9% = H9% + 1:H1% = CX - H2%: GOSUB 630: PRINT
: PRINT "CODE 0": HTAB 10: PRINT H1%: HTAB 20: PRINT "CODE
1": HTAB 29: PRINT H2%: PRINT : PRINT "AND. DES RUNS": HTAB
25: PRINT H9%: PRINT : PRINT "TESTWERT 2": HTAB 25: IF CX
= 30 THEN PRINT "-": GOTO 17100
17030 H1 = 2 * H1% * H2%:H3 = H1 / CX + 1:H4 = 500 * (H1 + H1
- CX) / CX / CX / (CX - 1): IF H4 = 0 THEN 740
17040 Z = (H2% - H3) / H4: IF CX < 50 THEN Z = Z + 500
/ (Z) / H4
17050 S = FN R( FN STN(Z)): PRINT FN R(Z)
17060 PRINT : PRINT "SIGNIFIKANZ": HTAB 25: PRINT S: GOSUB
650: IF H4% < > 0 THEN PRINT : INVERSE : PRINT "ACHTUNG":
NORMAL : PRINT : PRINT "ES TRETEN SIGN-WECHSEL AUF": PRINT "INNERHALB
VON TIE-GRUPPEN.": GOSUB 700: GOTO 1000
17100 H9 = CX:H8 = H1%: GOSUB 310:Q = H1: IF H9% = 2 THEN S
= FN R(1 / Q): GOTO 17060
17110 H9 = H9% / 2:H3% = 1: IF INT (H9) < H8 THEN H8 = (H9%
+ 1) / 2:H3% = 0
17120 H8% = H8 - 1:H6% = H1% - 1:H7% = H2% - 1: FOR J = 1 TO
H8%:H9 = H4%:H8 = J - 1: GOSUB 310:H6 = H1:H9 = H7%:H5 = J -
1: GOSUB 310:H7 = H1:S = S + 2 * H6 * H7
17130 H8 = J: GOSUB 310:S = S + H6 * H1:H6 = H1:H9 = H6%:H5
= J: GOSUB 310:S = S + H7 * H1: NEXT :S = S + 2 + H6 * H1 *
H3%:S = FN R(S / Q): GOTO 17060

```

```

17000 A# = "MCNEMAR": GOSUB 600:VC% = 2: GOSUB 420:DUX = C%
/ 2.7: GOSUB 520: GOSUB 670: GOSUB 660: GOSUB 720: IF C% = 0
THEN 710
18010 FOR J = 1 TO C%:A(0,J) = Y(J):X(J) = X(J):NEXT J: GOSUB
720: GOSUB 610: FOR J = 1 TO C%:Y = A(0,J):X = X(J):H3% =
N%(X,Y) + 1: NEXT J:H3% = N%(0,1) + N%(1,0): IF H3% > 10 THEN
CH = (ABS(N%(0,1) - N%(1,0)) - 1) / 2: H3% = 1: GOSUB 260:
GOTO 18050
18030 K = N%(0,1): IF N%(1,0) < K THEN K = N%(1,0)
18040 H2 = H3%:P = .5: GOSUB 290:S = FN R(2 + B)
18050 GOSUB 630: PRINT: HTAB 17: PRINT V$(VC%:VM%(1)): HTAB
13: PRINT "0":; HTAB 21: PRINT "1": GOSUB 18500: HTAB 7: PRINT
"0":; HTAB 9: PRINT "I":; HTAB 12: PRINT N%(0,0):; HTAB 17:
PRINT "I":; HTAB 20: PRINT N%(1,0):; HTAB 25: PRINT "I": PRINT
V$(VC%(VM%(2))): GOSUB 18500
18060 HTAB 7: PRINT "1":; HTAB 9: PRINT "I":; HTAB 12: PRINT
N%(0,1):; HTAB 17: PRINT "I":; HTAB 20: PRINT N%(1,0):; HTAB
25: PRINT "I": GOSUB 18500: PRINT: PRINT: IF H3% > 10 THEN
PRINT "TESTWERT CH":; HTAB 25: PRINT FN R(CH): PRINT
18070 PRINT "SIGNIFIKANZ ":; HTAB 25: PRINT FN R(S): PRINT
GOSUB 650: GOSUB 700: GOTO 1800
18500 HTAB 9: PRINT "I-----I-----I": RETURN
18999 REM

```

```

19000 A# = "COCHRAN-Q-TEST": GOSUB 600: INPUT "ANZAHL DER GRUPPEN
? ":VC%: IF VC% > 5 THEN CALL - 198: VTAB 10: PRINT "ZU VIELE
GRUPPEN. ABRUCH": GOSUB 700: GOTO 1000
19010 GOSUB 420: GOSUB 560: GOSUB 670:DUX = C% / 2.5: FOR J
= 1 TO C%:X(J) = 0: NEXT J: GOSUB 660: IF C% = 0 THEN 710
19020 GOSUB 610: FOR J = 1 TO C%: FOR I = 1 TO VC%:X = A(VM%(I),J):
IF X < > 0 AND X < > 1 THEN HOME: VTAB 10: PRINT "VARIABLE
NICHT 0/1 CODIERT. ABRUCH": GOSUB 700: GOTO 1000
19030 TX(I) = TX(I) + X:X(J) = X(J) + X: NEXT J:P1% = H1% + X(J):H3%
= H2% + X(J) * X(J): NEXT J: FOR I = 1 TO VC%:H3% = H3% + TX(I):H4%
= H4% + TX(I) * TX(I): NEXT I:V = VC% - 1:H1 = VC% * H1% - H2%:
IF H1 < = 0 THEN S = IC%:CH = S: GOTO 19050
19040 CH = V * (VC% * H4% - H3% * H3%) / H1: GOSUB 260
19050 GOSUB 640: POKE 33,20: POKE 32,20: HOME: PRINT "CODE
0":; HTAB 10: PRINT "CODE 1": PRINT: FOR I = 1 TO VC%: HTAB
2: PRINT C% - TX(I):; HTAB 11: PRINT TX(I): NEXT I: PRINT: PRINT
: POKE 33,40: POKE 32,0: PRINT: PRINT: PRINT "FREIHEITSGRADE":
19060 HTAB 25: PRINT VC% - 1: PRINT: PRINT "TESTWERT Q":;
HTAB 25: PRINT FN R(CH): PRINT: PRINT "SIGNIFIKANZ":; HTAB
25: PRINT S: GOSUB 650:N%(0,0) = 0: FOR I = 1 TO S:N%(I) = 0: NEXT
I: GOSUB 700: GOTO 1000

```

```

20000 A# = "RUNS"; GOSUB 600:VC% = 1: GOSUB 420: GOSUB 540:
HOME : PRINT "SOLL AM MEDIAN DICHOTOMISIERT WERDEN ?": IF C%
= 0 THEN 710
20010 HTAB 20: INPUT "EINGABE (J/N) : ";A#: IF A# < > "J"
AND A# < > "N" THEN CALL - 998: CALL - 958: GOTO 20010
20020 GOSUB 660: GOSUB 670:DU% = CX / 6: GOSUB 610: IF A# =
"N" THEN FOR J = 1 TO CX:H7 = H7 + 1/J: NEXT :H7 = H7 / C%:
GOTO 20040
20030 FOR J = 1 TO CX:A(0,J) = X(J): NEXT : GOSUB 880:H7 =
H9: FOR J = 1 TO CX:X(J) = A(0,J): NEXT
20040 H3% = 1:H1 = H7 - X(1):H1% = SGN ( SGN (H1) + 1):TX(H1%)
= TX(H1%) + 1: FOR J = 2 TO CX:H2 = H7 - X(J):H2% = SGN ( SGN
(H2) + 1):TX(H2%) = TX(H2%) + 1: IF H1% < > H2% THEN H3% =
H3% + 1
20050 H1% = H2%: NEXT :H6% = 2 * TX(0) * TX(1):H3 = H6% / CX
+ 1:H4 = SQR (H6% * (H6% - CX) / CX / CX / (CX - 1)): IF H+
= 0 THEN 740
20060 Z = (H3% - H3) / H4: IF C% < 50 THEN Z = Z - SGN (Z)
* .5 / H4
20070 S = FN R(2 * FN STN(2)): GOSUB 620: PRINT : PRINT :
PRINT "ANZ.DER RUNS";: HTAB 25: PRINT H3%: PRINT : PRINT "ANZ.
DER + DIFFERENZEN";: HTAB 25: PRINT TX(1): PRINT : PRINT "ANZ.
DER - DIFFERENZEN";: HTAB 25: PRINT TX(0): PRINT
20080 PRINT "TESTWERT Z";: HTAB 25: PRINT FN R(Z): PRINT :
PRINT "SIGNIFIKANZ";: HTAB 25: PRINT S: PRINT : GOSUB 650: GOSUB
700: GOTO 1000

```



```

22000 A$ = "SPEARMAN": GOSUB 600:VC% = 2: GOSUB 420: GOSUB 520:
HOME : IF CX = 0 THEN 710
22010 GOSUB 670:DU% = CX / 2: GOSUB 610:N = CX: GOSUB 100:
GOSUB 180:H9 = H3:H9% = H4%: FOR J = 1 TO CX:A(0,J) = X(J): Y(J)
= Y(J):Y(J) = A(0,J): NEXT : GOSUB 100: GOSUB 180: FOR J = 1
TO CX:A(0,J) = X(J):X(J) = Y(J):Y(J) = A(0,J): NEXT
22020 FOR J = 1 TO CX:H8 = H8 + (X(J) - Y(J)) ^ 2: NEXT :H1
= N * N * N:H9 = (H1 - N - H9) / 12:H3 = (H1 - N - H3) / 12:
IF H9 * H3 = 0 THEN Z = 1CX:S = Z: GOTO 22040
22030 Z = (H9 + H3 - H8) / 2: SQR (H9 * H3):V = N - 2:T =
Z * SQR (ABS (V / (1 - Z * Z))) : GOSUB 210
22040 GOSUB 630: PRINT : PRINT : HTAB 15: PRINT "VARIABLE1
VARIABLE2": PRINT : PRINT "ANZ. DER TIES": HTAB 18: PRINT
H9%: HTAB 29: PRINT H4%: PRINT : PRINT : PRINT "KORRELATIONSKOEFF.":
HTAB 25: PRINT FN R(2): PRINT
22050 PRINT "FREIHEITSGRADE": HTAB 25: PRINT N - 2: PRINT
: PRINT "TESTWERT T": HTAB 25: PRINT FN R(T): PRINT : PRINT
"SIGNIFIKANZ": HTAB 25: PRINT S: PRINT : GOSUB 650: GOSUB 700:
GOTO 1000
22999 REM

```

```

23000 A$ = "BRAVAIS/PEARSON": GOSUB 600:VC% = 2: GOSUB 420:
GOSUB 520: HOME : IF CX = 0 THEN 710
23010 DU% = CX / 8: GOSUB 610: GOSUB 670: FOR J = 1 TO CX:
= X(J):Y = Y(J):H2 = H2 + X:H3 = H3 + Y:H4 = H4 + X * X:H5 =
H5 + Y * Y:H1 = H1 + X * Y: NEXT :H6 = (H4 - H2 * H2 / CX) -
(H5 - H3 * H3 / CX): IF H6 = 0 THEN 740
23020 Z = (H1 - H2 * H3 / CX) / SQR (H6):V = CX - 2:T = Z *
SQR ((CX - 2) / (1 - Z * Z)): GOSUB 210: GOSUB 630: PRINT :
PRINT : PRINT "KORRELATIONSKOEFF. R": HTAB 25: PRINT FN R(Z) :
PRINT
23030 PRINT "FREIHEITSGRADE": HTAB 25: PRINT V: PRINT : PRINT
"TESTWERT T": HTAB 25: PRINT FN R(T): PRINT : PRINT "SIGNIFIKANZ":
HTAB 25: PRINT S: PRINT : GOSUB 650: GOSUB 700: GOTO 1000

```

```

24000 A* = "MEDIANTEST": GOSUB 600: PRINT "HINWEIS: GRUPPIERUNG
DURCH VARIABLE 2": VC% = 2: GOSUB 700: HOME : GOSUB 420: GOSUB
520: HOME : IF C% = 0 THEN 710
24010 GOSUB 660: GOSUB 670: DU% = C% / 3: GOSUB 610: H1% = Y(1):
FOR J = 1 TO C%: Y = Y(J): IF Y < 3: INT(Y) THEN HOME : VTA=
10: CALL - 198: PRINT "FEHLER IN DER CODIERUNG": GOSUB 700:
GOTO 1000
24020 IF Y < H1% THEN H1% = Y
24030 IF Y > H2% THEN H2% = Y
24040 NEXT : H3% = H2% - H1%: IF H3% > 5 THEN HOME : VTA= 10:
CALL - 198: PRINT "ZUVIELE GRUPPEN. ABRUCH": GOSUB 700: GOTO
1000
24050 GOSUB 800: GOSUB 830: GOSUB 850: GOSUB 840: CH = 0: IF
H3% = 1 AND C% < = 30 THEN GOSUB 770: GOTO 24130
24060 IF H3% = 1 AND C% > 30 THEN 24100
24070 V = H3%: Q = 0: FOR I = 0 TO H3%: H8 = TX(0) * S(I) / N: H7
= TX(1) * S(I) / N: CH = CH + (NX(0,I) - H8) ^ 2 / H8 + (NX(1,I)
- H7) ^ 2 / H7: IF H8 < 5 THEN H8% = H8% + 1: IF H8 < 1 THEN
Q = 1
24080 IF H7 < 5 THEN H8% = H8% + 1: IF H7 < 1 THEN Q = 1
24090 NEXT : GOSUB 260: H5 = H8% ^ (2 * H3% + 1) / 5: ON SGN
(H5 + Q) GOSUB 24200: GOTO 24120
24100 V = 1: H8 = TX(0) * TX(1) + S(0) * S(1): IF H6 = 0 THEN
PRINT "TESTWERT NICHT ZU BERECHNEN": PRINT "NENNER IST NULL.
ABBRUCH": GOSUB 700: GOTO 1000
24110 CH = (ABS (NX(0,0) * NX(1,1) - NX(0,1) * NX(1,0)) - N
/ 2) ^ 2 * N / H6: GOSUB 260
24120 PRINT "FREIHEITSGRADE":; HTAB 25: PRINT H3%: PRINT :
PRINT "TESTWERT CHI":; HTAB 25: PRINT FN PCHI : PRINT : PRINT
"SIGNIFIKANZ":; HTAB 25: PRINT FN R(5)
24130 GOSUB 650: GOSUB 700: GOTO 1000
24200 POKE 34,15: HOME : PRINT : INVERSE : PRINT "ACHTUNG !!":
PRINT : PRINT : IF Q = 1 THEN PRINT "NICHT ALLE ERWARTETEN
H. SIND >= 1.": PRINT
24210 IF H5 = 1 THEN PRINT "MEHR ALS 20% DER ERWARTETEN H.
< 5."
24220 GOSUB 700: NORMAL : HOME : RETURN

```

```

25000 A$ = "MANN/WHITNEY-U-TEST": GOSUB 600: PRINT "HINWEIS: VARIABLE2
MUSS 0/1 CODIERT SEIN": UCX = 2: GOSUB 700: HOME : GOSUB 420:
GOSUB 520: HOME : IF CX = 0 THEN 710
25010 GOSUB 720: DUX = CX / 2.5: GOSUB 610
25020 N = CX: GOSUB 660: GOSUB 670: GOSUB 180: GOSUB 180: FOR
J = 1 TO CX: H8% = Y(J): S(H8%) = S(H8%) + X(J): TX(H8%) = TX(H8%)
+ 1: NEXT J: H7 = TX(0) * TX(1): H1 = H7 + TX(0) * (TX(0) + 1)
/ 2 - S(0): IF H1 > H7 / 2 THEN H1 = H7 - H1
25030 Z = (H1 - H7 / 2) / SQR (H7 - CX / 12 * (CX - 1) * (CX + 1)
* CX - CX - H3) / 12): S = FN R(2 * FN STN(Z)): GOSUB 630:
PRINT : P1% = 1: GOSUB 750
25040 PRINT : PRINT "U " : H1: HTAB 15: PRINT "H2. DER 3.03
": H4%: PRINT : PRINT "TESTWERT Z": HTAB 25: PRINT FN R(2):
PRINT : PRINT "SIGNIFIKANZ": HTAB 25: PRINT S: PRINT : GOSUB
650: GOSUB 700: GOTO 1000
25999 REM

```

```

26000 A$ = "KRUSKAL/WALLIS-H-TEST": GOSUB 600: PRINT "HINWEIS:
VARIABLE2 MUSS GANZZAHLLIG.": PRINT "VON 0 BIS MAXIMAL 5 CODIERT
SEIN.": UCX = 2: GOSUB 700: HOME : GOSUB 420: GOSUB 520: HOME
: IF CX = 0 THEN 710
26010 GOSUB 670: GOSUB 660: DUX = CX / 3.5: GOSUB 610: N = CX:
GOSUB 180: GOSUB 180: FOR J = 1 TO CX: X = X(J): Y = Y(J): S(Y)
= S(Y) + X: TX(Y) = TX(Y) + 1: IF Y > H1% THEN H1% = Y
26020 NEXT J: H9 = H3: CH = 0: FOR I = 0 TO H1%: IF S(I) > 0 THEN
H2% = H2% + 1: CH = CH + S(I) * S(I) / TX(I)
26030 NEXT I: H2% = H2% - 1: CH = CH * 12 / N * (N + 1) - 3 *
(N + 1): CH = CH / (1 - H9 + (N * N * N - N)) / 12 = H2%: GOSUB
260
26040 GOSUB 630: PRINT : GOSUB 750: PRINT "FREIHEITSGRADE":
HTAB 25: PRINT H2%: HTAB 20: PRINT "TIES. ": H4%: PRINT : PRINT
"TESTWERT CHI": HTAB 25: PRINT FN R(CH): PRINT
26050 PRINT "SIGNIFIKANZ": HTAB 25: PRINT S: GOSUB 650: GOSUB
700: GOTO 1000

```

```

27000 A$ = "DIXON/MOOD-VORZEICHEN-TEST": GOSUB 600:VC% = 2:
GOSUB 420: GOSUB 520: HOME : IF C% = 0 THEN 710
27010 GOSUB 670:DU% = C% / 8: GOSUB 610: FOR J = 1 TO C%:H9%
= SGN (X(J) - Y(J)): IF H9% = 1 THEN H1% = H1% + 1: GOTO 27030
27020 H2% = H2% + ABS (H9%)
27030 NEXT :H3% = H1% + H2%:K = H1%: IF H2% > H1% THEN K =
H2%
27040 IF H3% <= 25 THEN H9 = H3%:P = .5: GOSUB 290:3 = FN
R(2 * S):H7% = 1: GOTO 27050
27050 K = H3% - K:Z = (K / .5 - H3% - 1) / SQR (H3%):S = FN
R(2 * FN STN(Z))
27060 GOSUB 630: VTAB 9: PRINT "ANZ. DER + DIFFERENZEN": HTAB
25: PRINT H1%: PRINT : PRINT "ANZ. DER - DIFFERENZEN": HTAB
25: PRINT H2%: PRINT : PRINT "TESTWERT Z": HTAB 25: IF H7%
= 1 THEN PRINT "-": GOTO 27070
27065 PRINT FN R(Z)
27070 PRINT : PRINT "SIGNIFIKANZ": HTAB 25: PRINT S: PRINT
: GOSUB 650: GOSUB 700: GOTO 1000
27999 REM

```

```

28000 A$ = "WILCOXON-TEST FUER VERBUNDENE PAARE": GOSUB 600:VC%
= 2: GOSUB 420: GOSUB 520: HOME : IF C% = 0 THEN 710
28010 GOSUB 660: GOSUB 670:DU% = C% / 4: GOSUB 610: FOR J =
1 TO C%:H1 = X(J) - Y(J): IF SGN (H1) = 0 THEN 28030
28020 H9% = H9% + 1:X(H9%) = ABS (H1):Y(H9%) = SGN (H1): IF
Y(H9%) = 1 THEN H1% = H1% + 1
28030 NEXT :N = H9%: GOSUB 100: GOSUB 120: FOR J = 1 TO H9%:S(1)
= S(1) + X(J) * SGN (Y(J) + 1):S(2) = S(2) + 1 * SGN (Y(J)
- 1): NEXT : IF H1% = 0 OR H1% = H9% THEN 1 = 10:Z = S: GOTO
28060
28040 H2% = H9% - H1%:H3 = S(0): IF S(1) < S(0) THEN H3 = S(1)
28050 Z = (H3 - N * (N + 1) / 4) / SQR (N * (N + 1) * (2 *
N + 1) / 24):S = FN R(2 * FN STN(Z))
28060 GOSUB 630: PRINT : PRINT : PRINT "ANZ. DER TIES": HTAB
25: PRINT H4%: PRINT : PRINT "+ DIFFERENZEN ":H1%: HTAB 19:
PRINT "MITTL. RANG ": FN R(S(1) / H1%): PRINT
28070 PRINT "- DIFFERENZEN ":H2%: HTAB 19: PRINT "MITTL. RANG
": FN R(S(0) / H2%): PRINT : PRINT "TESTWERT Z": HTAB 25: PRINT
FN R(Z): PRINT : PRINT "SIGNIFIKANZ": HTAB 25: PRINT S: PRINT
: GOSUB 650: GOSUB 700: GOTO 1000

```



```

29000 A$ = "FRIEDMAN": GOSUB 300: INPUT "ANZAHL DER GRUPPEN
? "; VC%: IF VC% > 5 THEN CALL - 198: VTAB 10: PRINT "ZU VIELE
GRUPPEN. ABRUCH": GOSUB 700: GOTO 1000
29010 GOSUB 420: GOSUB 660: GOSUB 670: P = 0: DUX = NF%: GOSUB
610: GOSUB 560: IF CX = 0 THEN 710
29020 FOR K = 1 TO CX: FOR I = 1 TO VC%: X(I) = A(VM%(I),K): Y(I)
= I: NEXT :N = VC%: GOSUB 100: GOSUB 180: FOR I = 1 TO VC%: H1%
= Y(I): S(H1%) = S(H1%) + X(I): NEXT
29030 NEXT K: FOR I = 1 TO VC%: P = P + S(I) * S(I): NEXT :CP
= 12 / CX / VC% / (VC% + 1) * P + 3 * CX * (VC% + 1): M = VC%
- 1: GOSUB 260
29040 GOSUB 640: POKE 33,20: POKE 32,20: HOME : PRINT "MITTLERE
RAENGE": PRINT : FOR I = 1 TO VC%: PRINT FN R(S(I) / CX): NEXT
: POKE 32,0: POKE 33,40: PRINT : PRINT : PRINT "FREIHEITSGRADE
": HTAB 25: PRINT VC% - 1: PRINT
29050 PRINT "TESTWERT CHI ": HTAB 25: PRINT FN R(CH): PRINT
: PRINT "SIGNIFIKANZ ": HTAB 25: PRINT S: GOSUB 650: NCX = 0:
FOR I = 1 TO S: VC%(I) = 0: NEXT : GOSUB 700: GOTO 1000
29999 REM

61000 IF PEEK (222) = 254 THEN HTAB 5: CALL - 198: INVERSE
: PRINT "FALSCH EINGABE. WIEDERHOLUNG": FOR Q = 1 TO 1500:
NEXT : NORMAL : CALL - 998: CALL - 998: CALL - 958: RESUME

61010 HOME : POKE 768,104: POKE 769,168: POKE 770,104: POKE
771,166: POKE 772,223: POKE 773,154: POKE 774,72: POKE 775,152:
POKE 776,72: POKE 777,96: CALL 768
61020 HOME : IF PEEK (222) = 5 THEN VTAB 10: CALL - 198:
INVERSE : HTAB 12: PRINT "FALSCH DISKETTE": NORMAL : GOTO 61030
61030 IF PEEK (222) = 6 THEN VTAB 10: CALL - 198: HTAB 12:
PRINT "FALSCH DISKETTE": GOTO 61060
61040 IF PEEK (222) = 8 THEN VTAB 10: HTAB 4: CALL - 198:
PRINT "LESEN/SCHREIBEN UNMOEGLICH. WIEDERHOLUNG": GOTO 61060
61050 VTAB 9: CALL - 958: VTAB 10: HTAB 12: PRINT "FATAL ERROR":
PRINT : HTAB 7: PRINT "NEUSTART DES SYSTEMS": VTAB 22: END
61060 GOSUB 700: CLEAR : GOTO 10

```


Programme II : TESTS MIT HILFEN

```

*****
*
*          ***REP-TEST***
*
*   NICHT-PARAMETRISCHE TESTVERFAHREN
*
*   (C) RENATE PRUST 1983
*
*   TESTS MIT HILFEN
*
*****

```

```

10  DIM IO(201),A(5,201),Y(201),Y(201): DIM U$(82),UC%(5),UM%(5):
DIM TX(9),S(9),NX(9,9): DEF FN STN(Z) = EXP ( - ((83 + ABS
(Z) + 351) * ABS (Z) + 562) * ABS (Z) / (703 + 165 * ABS
(Z))) / 2:PI = 3.141592653: ONERR GOTO 61000
20  DEF FN R(Z) = INT (Z * 1000 + .5) / 1000:IC% = 999:F2$
= " VARDATA FILE ": HOME : CALL - 198: VTAB 10: HTAB 5: INVERSE
: PRINT "LEGEN SIE DIE DATENDISKETTE EIN": PRINT : HTAB 3: PRINT
" UND DRUECKEN SIE DIE RETURN-TASTE": NORMAL : GET A$: PRINT
: HOME
30  VTAB 10: HTAB 10:A$ = "VARCATALOG": PRINT "READING...":A$:
PRINT : HTAB 8: PRINT "DAUER: MAX. 15 SEKUNDEN": PRINT CHR$
(4):"OPEN":A$: PRINT CHR$ (4):"READ":A$: INPUT F$,NV%,NF%,MW,IS%,INX:
FOR J = 1 TO NV%: INPUT U$(J): NEXT : PRINT CHR$ (4):"CLOSE":A$:
GOTO 2000
100 H1 = N - 1
110 H1 = INT ((H1 - 2) / 3) + 1: FOR H2 = 1 TO N - H1:H3 =
H2 + H1:H4 = X(H3):H5 = Y(H3): IF X(H2) < = H4 THEN 160
120 H6 = H2
130 X(H3) = X(H6):Y(H3) = Y(H6):H3 = H6: IF H6 < = H1 THEN
150
140 H6 = H6 - H1: IF X(H6) > H4 THEN 130
150 X(H3) = H4:Y(H3) = H5
160 NEXT H2: IF H1 > 1 THEN 110
170 RETURN

```

```

180 H4% = 0:H2 = 0:H3 = 0: IF X(1) = X(N) THEN FOR J = 1 TO
N:X(J) = (N + 1) / 2: NEXT :H4 = N * N:H3 = H4 * N - N:H2 =
H4 - N:H4% = N: RETURN
190 X(N + 1) = X(1):S = 1: FOR J = 1 TO N: IF X(J) = X(J + 1)
THEN NEXT
200 H1 = (S + J) / 2: FOR Q = S TO J:X(Q) = H1: NEXT :H4 = J
- S + 1:H1 = H4 * H4:H3 = H3 + H1 * H4 - H4:H2 = H1 - H4:H4%
= H4% + H4 * SGN (H4 - 1):S = J + 1: NEXT : RETURN
210 S = 0: IF V > 30 THEN S = FN R(2 * FN STN(T - T / 4
V) / SQR (1 + T * T / 2 / V)): RETURN
220 T = ABS (T):H9 = 1:H8 = T / SQR (V):H7 = ATN (H8):H6
= H8 / SQR (1 + H8 * H8):H5 = H6 / H8:: IF V = 1 THEN S
= FN R(1 - 2 * H7 / PI): RETURN
230 H1% = V - 3:H2% = INT (V / 2) = V / 2:H3 = 2 - H2%:: IF
H3 < = H1% THEN FOR H1 = H3 TO H1% STEP 2:H9 = H9 * H1 - (H1
+ 1):S = S + H9 * H5 ^ (H1 + 1): NEXT
240 IF NOT H2% THEN S = FN R(1 - ((H5 + S) * H6 + H7) + 2
/ PI): RETURN
250 S = FN R(1 - H6 * S - H6): RETURN
260 IF CH > 90 OR CH < .0000001 OR V > 35 THEN S = IC%: RETURN

270 H1 = 1: FOR H2 = V TO 2 STEP - 2:H1 = H1 * H2: NEXT :H9
= CH ^ INT ((V + 1) / 2) * EXP (- CH / 2) / H1:H4 = 1: IF
INT (V / 2) < > V / 2 THEN H4 = SQR (2 / CH / PI)
280 H7 = .00000001:H5 = 1:H6 = 1: FOR Q = 0 TO 1:V = V + 2:H6
= H6 * CH / V:H5 = H5 + H6:Q = H6 < H7: NEXT :S = FN R(1 -
H4 * H3 * H5): RETURN
290 Q = 1 - P:H6 = 1:H7 = Q * H9:S = FN R(H6 * H7): IF * =
0 THEN RETURN
300 FOR J = 1 TO K:H8 = J: GOSUB 310:H6 = H6 * P:H7 = H7
Q:S = S + H1 * H6 * H7: NEXT :S = FN R(S): RETURN
310 H1 = 1: IF H8 = 0 OR H9 = H8 THEN RETURN
320 H2 = 1: IF H9 = H8 > H8 THEN H9 = H9 - H8
330 FOR H3 = 1 TO H9 - H8:H1 = H1 * H3: NEXT : FOR H3 = H8
+ 1 TO H9:H2 = H2 * H3: NEXT :H1 = H2 / H1: RETURN
400 UTAB 15: HTAB 10: PRINT "LOADING...VARIABLE": PRINT : HTAB
11: PRINT "DAUER: 7 SEKUNDEN"
410 A$ = V$(VC$(IX)) + F2$ + F$: PRINT CHR$(4);"ELOAD";A$;":A":
PEEK (107) + PEEK (108) + 256: FOR Q = 1 TO NF:A(1),Q) = IC(Q):
NEXT : RETURN
420 H1 = 5: FOR J = 1 TO VC%: UTAB (H1 + J): PRINT "VARIABLE
";J:" ": INPUT "":A$: GOSUB 440:VM$(J) = I%: UTAB (H1 + J):
HTAB 20: PRINT V$(VM$(J)): NEXT : FOR J = 1 TO VC%: GOSUB 470:
NEXT J: FOR I = 1 TO VC%: FOR J = 1 TO NC%: IF VM$(I) = VC$(J)
THEN VM$(I) = J
425 NEXT : NEXT : RETURN
430 HTAB 6: INPUT "GEWUENSCHTE VARIABLE ":A$
440 IF A$ = "" THEN POP : GOTO 1000
450 I% = 0: FOR I = 1 TO NV%: IF A$ < > V$(I) THEN NEXT :
UTAB 20: HTAB 5: PRINT "VARIABLE ";A$:: PRINT " EXISTIERE
NICHT": FOR Q = 1 TO 1500: NEXT : CALL - 998: CALL - 999:
GOTO 430
460 I% = 1: RETURN

```

```

470 FOR I = 1 TO NC%: IF VM%(J) = VC%(I) THEN RETURN
480 NEXT : IF NC% < 5 THEN NC% = NC% + 1:IX = NC%: GOTO 510
490 FOR I = 1 TO NC%: FOR K = 1 TO VC%: IF VM%(K) = VC%(I)
THEN NEXT I
500 NEXT :IX = I
510 VC%(IX) = VM%(J): GOSUB 400: RETURN
520 CX = 0: FOR J = 1 TO NF%: X = A(VM%(1),J): Y = A(VM%(2),J):
IF Y < > MW AND X < > MW THEN CX = CX + 1: X(CX) = X: Y(CX)
= Y
530 NEXT : RETURN
540 CX = 0: FOR J = 1 TO NF%: X = A(VM%(1),J): IF X < > MW THEN
CX = CX + 1: X(CX) = X
550 NEXT : RETURN
560 CX = 0: FOR J = 1 TO NF%: FOR I = 1 TO VC%: X = A(VM%(I),J):
IF X < > MW THEN NEXT : CX = CX + 1: FOR I = 1 TO VC%: A(VM%(I),CX)
= A(VM%(I),J): NEXT
570 NEXT J: RETURN
600 HOME : H1% = LEN (A$): H2% = INT ((40 - H1%) / 2): FOR
J = 1 TO H2%: PRINT "-": NEXT : PRINT A$: FOR J = H2% + H1%
+ 1 TO 40: PRINT "-": NEXT : POKE 34,4: HOME : RETURN
610 HOME : VTAB 10: HTAB 14: PRINT "RECHNUNG": PRINT : HTAB
8: PRINT "DAUER: MAX. "; DU% + 1: " SEKUNDEN": RETURN
620 HOME : VTAB 5: PRINT "VARIABLE "; V$(VC%(VM%(1))): RETURN

630 HOME : PRINT "VARIABLE "; V$(VC%(VM%(1))): " MIT VARIABLE
"; V$(VC%(VM%(2))): RETURN
640 HOME : PRINT "TESTVARIABLEN": PRINT : FOR I = 1 TO VC%:
PRINT V$(VC%(VM%(I))): NEXT : RETURN
650 PRINT : PRINT "GUELTIGE FAELLE "; CX: "/ FEHLENDE FAELLE
"; NF% - CX: RETURN
660 FOR J = 0 TO 9: S(J) = 0: T%(J) = 0: FOR I = 0 TO 9: N%(I,J)
= 0: NEXT : NEXT : RETURN
670 H1 = 0: H2 = 0: H3 = 0: H4 = 0: H5 = 0: H6 = 0: H7 = 0: H8 = 0: H9
= 0: H1% = 0: H2% = 0: H3% = 0: H4% = 0: H6% = 0: H7% = 0: H8% = 0: H9%
= 0: RETURN
700 VTAB 24: HTAB 6: INVERSE : PRINT "DRUECKEN SIE IRGEND EINE
TASTE": NORMAL : POKE - 19368,0: WAIT - 16384,128: POKE
- 16368,0: PRINT : RETURN
710 HOME : VTAB 10: CALL - 198: PRINT "KEINE GUELTIGEN FAELLE.
ABBRUCH": GOSUB 700: GOTO 1000
720 FOR J = 1 TO CX: H1 = Y(J): IF H1 = 0 OR H1 = 1 THEN NEXT
: RETURN
730 HOME : VTAB 10: CALL - 198: PRINT "FEHLER: VARIABLE NICHT
0/1 CODIERT.": POP : GOSUB 700: GOTO 1000
740 HOME : VTAB 8: CALL - 198: PRINT "STANDARDABWEICHUNG =
0. ABRUCH ": GOSUB 700: GOTO 1000
750 PRINT "CODE ANZAHL MITTL. RANG": PRINT : FOR I = 0 TO
H1%: IF T%(I) < > 0 THEN HTAB 2: PRINT I: HTAB 9: PRINT T%(I):
HTAB 16: PRINT FN R(S(I) / T%(I))
760 NEXT : PRINT : RETURN

```

```

770 S = 0: IF NX(0,1) * NX(1,0) > NX(0,0) * NX(1,1) THEN X =
NX(0,0):NX(0,0) = NX(1,0):NX(1,0) = NX(1,1):NX(1,1) = NX(0,1):NX(0,1)
= X
780 H9 = NX(0,0) + NX(1,0):H8 = NX(0,0): GOSUB 310:Q = H1:H9
= CX - H9:H8 = NX(0,1): GOSUB 310:Q = Q * H1:H9 = CX:H8 = NX(0,0)
+ NX(0,1): GOSUB 310:S = S + Q / H1: IF NX(0,1) > 0 AND NX(1,0)
> 0 THEN NX(0,0) = NX(0,0) + 1:NX(0,1) = NX(0,1) - 1:NX(1,0)
= NX(1,0) - 1:NX(1,1) = NX(1,1) + 1: GOTO 780
790 PRINT "FISHER'S WAHRSCH.": HTAB 25: PRINT FN R(S): PRINT
: RETURN
800 N = CX: GOSUB 100:H6 = N / 2:H9% = H6 + .5:H9 = X(H9%):
IF INT (H6) = H6 THEN H9 = (X(H6) + X(H6 + 1)) / 2
810 FOR J = 1 TO CX:X = 0: IF X(J) > H9 THEN X = 1
820 X(J) = X: NEXT : RETURN
830 FOR J = 1 TO CX:Y = Y(J) - H1:X = X(J):NX(X,Y) = NX(X,Y)
+ 1:TX(X) = TX(X) + 1:S(Y) = S(Y) + 1: NEXT : GOSUB 630: PRINT
: RETURN
840 FOR I = 0 TO H3%: HTAB 15: PRINT "CODE ";I + H1%: HTAB
26: PRINT NX(0,I): HTAB 31: PRINT NX(1,I): HTAB 36: PRINT
S(I): NEXT : HTAB 26: PRINT TX(0): HTAB 31: PRINT TX(1): HTAB
36: PRINT CX: PRINT : RETURN
850 PRINT "MEDIAN: ": FN R(H9): HTAB 25: PRINT "<=< >=":
RETURN
860 H3 = Y(1) - X(1):H4 = H3: FOR J = 2 TO N:H2 = Y(J) - X(J):
IF H2 > H3 THEN H3 = H2
870 IF H2 < H4 THEN H4 = H2
880 IF A% = 3 THEN 910
890 H1 = Y(J - 1) - X(J): IF H1 > H3 THEN H3 = H1
900 IF H1 < H4 THEN H4 = H1
910 NEXT :H5 = ABS (H3): IF ABS (H4) > H5 THEN H5 = ABS
(H4)
920 S = 0:Z = H5 * SQR (V): IF Z < .27 THEN S = 1
930 IF Z > = .27 AND Z < 1 THEN Q = EXP ( - 1.233701 / Z
/ Z):H7 = Q ^ 9:S = 1 - 2.506628 / Z * (Q + H7 + H7 * H7 * H7
/ Q / Q)
940 IF Z > = 1 AND Z < 3.1 THEN Q = EXP ( - 2 * Z * Z):H7
= Q ^ 4:S = 2 * (Q - H7 + H7 * H7 / Q - H7 * H7 * H7 * H7)
950 RETURN
960 HTAB 6: PRINT "ABS. DIFF. POS. DIFF. NEG. DIFF.": PRINT
: PRINT "MAX.": HTAB 8: PRINT FN R(H5): HTAB 20:A$ = "-":
IF H3 > = 0 THEN A$ = STR$ ( FN R(H3))
970 PRINT A$: HTAB 32:A$ = "-": IF H4 < 0 THEN A$ = STR$
( FN R(H4))
980 PRINT A$: PRINT : PRINT : PRINT "TESTWERT Z": HTAB 25:
PRINT FN R(Z): PRINT : PRINT "SIGNIFIKANZ": HTAB 25: PRINT
FN R(S): RETURN

```

```

1000 HOME : VTAB 10: PRINT "WOLLEN SIE UNTERSCHIEDE": PRINT
: PRINT "IN DEN ";A$: PRINT : PRINT "PRUEFEN ?"
1010 VTAB 17: HTAB 10: INPUT "EINGABE (J/N) : ";A$: IF A$ <
> "J" AND A$ < > "N" THEN CALL - 998: CALL - 958: GOTO
1010
1020 RETURN
1030 HOME : PRINT "ES IST KEIN TEST AUF MITTELWERTUNTER-":
PRINT : PRINT "SCHIEDE VERFUEGBAR, DA DIE AUS: PRINT : PRINT
"MINIALEM SKALENNIVEAU ARBEITEN.": PRINT : PRINT "WOLLEN SIE
EINEN ANPASSUNGSTEST ?"
1040 INPUT A$: IF A$ < > "J" AND A$ < > "N" THEN CALL -
998: CALL - 958: GOTO 1040
1050 IF A$ = "J" THEN RETURN
1060 POF : GOTO 2000
1070 HOME : PRINT "WIEVIELE KATEGORIEN HAT DIE VARIABLE ?":
PRINT
1080 HTAB 25: INPUT "EINGABE: ";A$: IF A$ < > STR$ ( VAL
(A$)) OR VAL (A$) < 2 THEN CALL - 998: CALL - 958: GOTO
1080
1090 RETURN
1100 VTAB 9: PRINT "UNTERSCHIEDUNG,": PRINT "RANGORDNUNG UND":
PRINT "ABSTAEDE ERKENNBAR": VTAB 12: PRINT "UNTERSCHIEDUNG
UND": PRINT "RANGORDNUNG MOEGL.": VTAB 15: PRINT "NUR UNTERSCHIEDUNG":
PRINT "MOEGlich": RETURN
1999 REM

2000 YY = FRE (0):P% = 0:M% = 0:S% = 0:PS% = 0:R% = 0: HOME
: HTAB 11: PRINT "AUSWAHL MIT HILFEN": POKE 32,0: POKE 33,37:
PRINT : PRINT : HTAB 6: PRINT "WAS SOLL GEPRUEFT WERDEN?": PRINT
: PRINT
2010 PRINT "ABWEICHUNG EINER EMPIRISCHEN VER-": PRINT "TEILUNG
VON EINER THEORETISCHEN...1": PRINT : PRINT "ZUFAELLIGKEIT EINER
STICHPROBE....2": PRINT : PRINT "UNTERSCHIEDE ZWISCHEN GRUPPEN....3":
PRINT
2020 PRINT "UNTERSCHIEDE ZWISCHEN MESSWERT-": PRINT "REIHEN
DER GLEICHEN GRUPPE.....4": PRINT : PRINT "ZUSAMMENHANG ZWISCHEN
MERKMALEN...5": PRINT : PRINT "REP-TEST MENU.....6":
PRINT
2030 INPUT "AUSWAHL.....":P%: IF P% <
1 OR P% > 6 THEN CALL - 998: CALL - 958: GOTO 2030
2040 POKE 32,0: POKE 33,40: IF P% = 6 THEN HOME : VTAB 10:
HTAB 3: INVERSE : PRINT "LEGEN SIE DIE SYSTEMDISKETTE EIN":
PRINT : HTAB 3: PRINT "UND DRUECKEN SIE DIE RETURN-TASTE": NORMAL
: INPUT "":A$: HOME : VTAB 12: HTAB 12: PRINT "ZURUECK ZUM MENU":
PRINT CHR$ (4);"RUN REP-TEST MENU"
2050 HOME : HTAB 3: PRINT "WELCHE DATENQUALITAET LIEGT VOR
?": VTAB 9: PRINT "METRISCHE DATEN.1": VTAB 12: PRINT "ORDINALDATEN...
.2": VTAB 15: PRINT "NOMINALDATEN....3"

```



```

2060  VTAB 21: PRINT "HILFE ERHALTEN SIE DURCH DIE H-TASTE":
POKE 32,19: POKE 33,20: POKE 35,18: POKE 34,4: HOME
2070  VTAB 5: HTAB 6: INPUT "EINGABE: ";A$: IF A$ < > STR$
( VAL (A$)) AND A$ < > "H" THEN CALL - 998: CALL - 958:
GOTO 2070
2080  IF A$ = "H" THEN HOME : GOSUB 1100: GOTO 2070
2090  M% = VAL (A$): IF M% < 1 OR M% > 3 THEN CALL - 998:
CALL - 958: GOTO 2070
2100  POKE 35,23: POKE 34,2: POKE 33,40: POKE 32,0: HOME : IF
P% < > 3 AND P% < > 4 THEN 2210
2110  IF P% = 3 THEN A$ = "STICHPROBEN BZW. GRUPPEN,"
2120  IF P% = 4 THEN A$ = "MESSUNGEN PRO FALL,"
2130  PRINT "ANZAHL DER ";A$: PRINT : PRINT "DIE VERGlichen
WERDEN SOLLen ?"
2140  VTAB 10: HTAB 20: INPUT "EINGABE: ";A$: IF A$ < > STR$
( VAL (A$)) OR VAL (A$) < 2 THEN CALL - 998: CALL - 958:
GOTO 2140
2150  S% = VAL (A$)
2160  A$ = "MITTELWERTEN": GOSUB 1000: IF A$ = "J" THEN PS% =
2: GOTO 2210
2170  A$ = "HAEUFIGKEITSVERTEILUNGEN": GOSUB 1000: IF A$ = "J"
THEN PS% = 1: GOTO 2210
2180  HOME : VTAB 10: PRINT "MEHR MOEGlichKEITEN HABEN SIE NICHT.":
PRINT : PRINT "SIE KOENNEN JETZT ALLE ANGEBOte NOCH": PRINT
: PRINT "EINMAL DURCHGEHEN.": PRINT : PRINT "FALLS SIE ABRECHEN
WOLLEN, DRUECKEN": PRINT
2190  PRINT "SIE EINFACH <RETURN>": PRINT : PRINT "SONST IRGendeINE
ANDERE TASTE ": GET A$: IF A$ = CHR$ (13) THEN 2000
2200  GOTO 2160
2210  ON P% - 1 GOTO 2250,2270,2330,2360
2220  IF M% = 1 THEN 13000
2230  GOSUB 1070:R% = VAL (A$): IF M% = 2 THEN ON SGN ( SGN
(4 - R%) + 1) + 1 GOTO 13000,12000
2240  ON SGN (R% - 2) + 1 GOTO 11000,12000
2250  IF M% < 3 THEN 14000
2260  CALL - 198: VTAB 8: PRINT "ES STEHT KEIN TEST ZUR VERFUEGUNG,":
PRINT : PRINT "DA DER MEDIAN NICHT BESTIMMT": PRINT : PRINT
"WERDEN KANN.": GOSUB 700: GOTO 2000
2270  IF PS% = 2 THEN 2310
2280  IF S% > 2 THEN 15000
2290  IF M% < 3 THEN 16000
2300  GOSUB 1070:R% = VAL (A$): ON SGN (R% - 2) + 1 THEN 17000,15000
2310  IF M% = 3 THEN GOSUB 1030: GOTO 2280
2320  ON SGN (S% - 2) + 1 GOTO 18000,19000
2330  IF PS% = 1 THEN ON SGN (S% - 2) + 1 GOTO 20000,21000
2340  IF M% = 3 THEN GOSUB 1030:PS% = 1: GOTO 2330
2350  ON SGN (S% - 2) + 1 GOTO 20000,23000
2360  ON M% GOTO 24000,25000,26000

```

```

11000 A$ = "BINOMIAL": GOSUB 600:VCX = 1: GOSUB 420: GOSUB 540:
HOME : IF CX = 0 THEN 710
11010 FOR J = 1 TO CX:Y(J) = X(J): NEXT : GOSUB 720: VTAB 7:
PRINT "WAHRSCH. CODE "A$": HTAB 26: INVERSE : PRINT "VOREINST.=0.5":
NORMAL
11020 HTAB 18: INPUT A$:P = .5: IF A$ < "J" THEN P = .1
(A$): IF P < 0 OR P > 1 THEN CALL - 998: CALL - 958: GOTO
11030
11030 DUX = NFX / 16: GOSUB 610:H9% = 0: FOR J = 1 TO CX:H9%
= H9% + X(J): NEXT : IF CX < 26 THEN H9 = CX: H9% = H9: GOSUB
290: GOTO 11050
11040 Z = (H9% - SGN (H9% - CX * P) * .5 - CX * P) / SQR (CX
* P * (1 - P)):S = FN R1(FN SIN(Z))
11050 GOSUB 620: VTAB 8: PRINT "CODE HAEUF. REL. WAHRSCH.":
PRINT : PRINT " 0": HTAB 10: PRINT CX - H9%: HTAB 18: PRINT
FN R1 - H9% / CX: HTAB 26: PRINT FN R1 - S: PRINT : PRINT
" 1": HTAB 10: PRINT H9%: HTAB 18: PRINT FN R1H9% / CX:
HTAB 26: PRINT S: PRINT : PRINT
11060 PRINT "BERNOULLI-PARAMETER ":P: PRINT : GOSUB 650: GOSUB
700: GOTO 1000
11999 REM

```

```

12000 A$ = "CHI QU.-ANPASSUNGSTEST": GOSUB 600:VCX = 1: GOSUB
420: GOSUB 540: IF CX = 0 THEN 710
12010 GOSUB 670:DUX = CX / 10 + 1: GOSUB 680:H1% = X(1):H2%
= X(1): FOR J = 2 TO CX: X = X(J): IF H1% > X THEN H1% = X
12020 IF H2% < X THEN H2% = X
12030 NEXT :H3% = H2% - H1% + 1: IF H3% > 8 THEN VTAB 18:
PRINT "ZU VIELE KATEGORIEN. ABRUCH": GOSUB 700: GOTO 1000
12040 HOME : INPUT "PRUEFUNG AUF GLEICHVERTEILUNG ? (J/N) ":A$:
IF A$ < > "J" AND A$ < > "N" THEN CALL - 998: CALL - 958:
GOTO 12040
12050 FOR J = 1 TO CX:X = X(J) - H1%:T1(X) = T1(X) + 1: NEXT
: IF A$ = "J" THEN FOR I = 0 TO H3% - 1:S(I) = 0: H3%: NEXT
: GOTO 12070
12060 HOME : PRINT "CODE BEOBACHTET ERWARTET": PRINT
:Y = 0: FOR I = 0 TO H3% - 1: HTAB 2: PRINT I + H1%: HTAB 12:
PRINT T%(I): HTAB 23: INPUT S(I):Y = Y + S(I): NEXT : IF Y
< > CX THEN CALL - 198: PRINT : PRINT "SUMME DER ERWARTETEN
HAEUFIGK. ":Y: PRINT "ANZAHL DER GUELTIGEN FAELE ":CX: GOSUB
700: GOTO 12060
12070 GOSUB 610:CH = 0: FOR I = 0 TO H3% - 1:CH = CH + (T%(I)
- S(I)) ^ 2 / S(I): IF S(I) < 5 THEN HOME : PRINT "FEHLER:
E(I) < 5. ABRUCH": GOSUB 700: GOTO 1000
12080 NEXT :V = H3% - 1: GOSUB 260: GOSUB 620: PRINT : PRINT
"CODE BEOBACHTET ERWARTET": PRINT : FOR I = 0 TO H3% -
1: HTAB 2: PRINT I + H1%: HTAB 12: PRINT T%(I): HTAB 23: PRINT
FN R1S(I): NEXT : PRINT
12090 PRINT "FREIHEITSGRADE": HTAB 25: PRINT H3% - 1: PRINT
"TESTWERT CH": HTAB 25: PRINT FN R1CH: PRINT "SIGNIFIKANZ":
HTAB 25: PRINT S: GOSUB 650: GOSUB 700: GOTO 1000

```

```

13000 A$ = "KOLMOGOROFF/SMIRNOW-1 STICHPROBEN-TEST": GOSUB 600:VC%
= 1: GOSUB 420: GOSUB 540: HOME : IF C% = 0 THEN 710
13010 GOSUB 670:DU% = C% / 5: PRINT "WELCHE VERTEILUNG WIRD
ALS": PRINT : PRINT "THEORETISCHE VERTEILUNG ANGENOMMEN?": POKE
32,4: POKE 33,35: PRINT : PRINT : PRINT "GLEICHVERTEILUNG.....1":
PRINT : PRINT "NORMALVERTEILUNG.....2": PRINT : PRINT "POISSONVERTEILU
NG....3": PRINT
13020 INPUT "AUSWAHL.....":A%: IF A% < 1 OR A% > 3
THEN CALL - 998: CALL - 958: GOTO 13020
13030 POKE 32,0: POKE 33,40: GOSUB 610:N = C%: GOSUB 100:N
= 0:Y = X(1): FOR J = 1 TO C%:X = X(J):H9 = H9 + X:H8 = H8 +
X * X: IF X < > Y THEN N = N + 1:Y(N) = (J - 1) / C%:Y = X
13040 NEXT :H9 = H9 / C%:N = N + 1:Y(N) = 1: ON A% GOSUB 13100,13200,
13300: GOSUB 670:Y(0) = 0:V = C%: GOSUB 860: GOSUB 620: PRINT
: PRINT : GOSUB 960: GOSUB 650: VTAB 5: HTAB 20: PRINT V$(0):
GOSUB 700: GOTO 1000
13100 DU% = C% / 10: HOME : VTAB 8: PRINT "WOLLEN SIE MINIMUM
UND": PRINT : PRINT "MAXIMUM ANGEBEN (J/N) ? ": INPUT "":A$:
IF A$ = "N" THEN H1 = X(1):H2 = X(C%): GOTO 13120
13110 PRINT : PRINT : INPUT "MINIMUM ? ":H1: PRINT : INPUT
"MAXIMUM ? ":H2
13120 GOSUB 610:Q = 0:Y = X(1) - 1: FOR J = 1 TO C%:X = X(J):
IF X < > Y THEN Q = Q + 1:X(Q) = (X - H1) / (H2 - H1):Y = X
13130 NEXT :V$(0) = "GLEICHVERT.": RETURN
13200 DU% = C% / 10: HOME : VTAB 8: PRINT "WOLLEN SIE PARAMETER
ANGEBEN?": PRINT : PRINT : INPUT "MITTELWERT ? ":A$: IF A$
< > "" THEN H9 = VAL (A$)
13210 PRINT : INPUT "STANDARDABWEICHUNG ? ":A$: IF A$ < >
"" THEN H7 = VAL (A$): GOTO 13230
13220 H7 = SQR ((H8 - H9 * H9 * C%) / (C% - 1)): IF H7 = 0
THEN POP : GOTO 740
13230 GOSUB 610:Q = 0:Y = X(1) - 1: FOR J = 1 TO C%:X = X(J):
IF X < > Y THEN Q = Q + 1:Z = (X - H9) / H7:X(Q) = FN STN(Z):Y
= X: IF Z > 0 THEN X(Q) = 1 - X(Q)
13240 NEXT :V$(0) = "NORMALVERT.": RETURN
13300 FOR J = 1 TO C%:X = X(J): IF X < > INT (X) OR X < 0
THEN HOME : VTAB 10: PRINT "DIE DATEN MUESSEN GANZZAHLIG UND":
PRINT "NICHT-NEGATIV SEIN. ABRUCH": POP : GOSUB 700: GOTO 1000
13310 DU% = C% / 10: HOME : VTAB 8: PRINT "WOLLEN SIE DEN MITTELWERT":
PRINT : INPUT "EINGEBEN ? (J/N) ":A$: IF A$ = "J" THEN PRINT
: PRINT : INPUT "MITTELWERT: ":H9
13320 GOSUB 610: IF H9 = 0 THEN HOME : VTAB 10: PRINT "MITTELWERT
= 0. ABRUCH": POP : GOSUB 700: GOTO 1000
13330 H1 = EXP ( - H9):Q = 1:H3 = 0:H4 = H3:X = X(Q): IF X
< > 0 THEN FOR I = 1 TO X:H3 = H3 * H9 / I:H4 = H4 + H3: NEXT
13340 X(Q) = H4 * H1:Y = X: FOR J = 2 TO C%:X = X(J): IF X <
> Y THEN FOR I = Y + 1 TO X:H3 = H3 * H9 / I:H4 = H4 + H3:
NEXT :Q = Q + 1:X(Q) = H4 * H1:Y = X
13350 NEXT :V$(0) = "POISSONVERT.": RETURN

```

```

14000 A$ = "RUNS": GOSUB 600:VC% = 1: GOSUB 420: GOSUB 540:
HOME : IF C% = 0 THEN 710
14020 GOSUB 660: GOSUB 670:DU% = C% / 6: GOSUB 610: IF M% =
1 THEN FOR J = 1 TO C%: H7 = H7 + X(J): NEXT : H7 = H7 / C%:
GOTO 14040
14030 FOR J = 1 TO C%: A(0,J) = X(J): NEXT : GOSUB 800: H7 =
H9: FOR J = 1 TO C%: X(J) = A(0,J): NEXT
14040 H3% = 1: H1 = H7 - X(1): H1% = SGN ( SGN (H1) + 1): T%(H1%)
= T%(H1%) + 1: FOR J = 2 TO C%: H2 = H7 - X(J): H2% = SGN ( SGN
(H2) + 1): T%(H2%) = T%(H2%) + 1: IF H1% < > H2% THEN H3% =
H3% + 1
14050 H1% = H2%: NEXT : H6% = 2 * T%(0) * T%(1): P3 = H6% / C%
+ 1: H4 = SQR (H6% * (H6% - C%) / C% / C% / (C% - 1)): IF H4
= 0 THEN 740
14060 Z = (H3% - H3) / H4: IF C% < 50 THEN Z = Z - SGN (Z)
* .5 / H4
14070 S = FN R(2 * FN STN(Z)): GOSUB 620: PRINT : PRINT :
PRINT "ANZ.DER RUNS":; HTAB 25: PRINT H3%: PRINT : PRINT "ANZ.
DER + DIFFERENZEN":; HTAB 25: PRINT T%(1): PRINT : PRINT "ANZ.
DER - DIFFERENZEN":; HTAB 25: PRINT T%(0): PRINT
14080 PRINT "TESTWERT Z":; HTAB 25: PRINT FN R(2): PRINT :
PRINT "SIGNIFIKANZ":; HTAB 25: PRINT S: PRINT : GOSUB 650: GOSUB
700: GOTO 1000
14999 REM

```

```

15000 A$ = "CHI QU.-HOMOGENITAETSTEST": GOSUB 600:VC% = 2: PRINT
"HINWEIS: DIE VARIABLEN MUESSEN": PRINT "GANZZAHLIG, VON 0 AUFWAERTS":
PRINT "BIS HOECHSTENS 9, CODIERT SEIN": GOSUB 700: HOME : GOSUB
420: GOSUB 520: HOME : IF C% = 0 THEN 710
15010 GOSUB 670: GOSUB 660:DU% = C% / 4: GOSUB 610: FOR J =
1 TO C%: X = X(J): Y = Y(J): IF X > 9 OR Y > 9 THEN PRINT "FALSCHER
CODIERUNG. ABRUCH": GOSUB 700: GOTO 1000
15020 N%(X,Y) = N%(X,Y) + 1: T%(X) = T%(X) + 1: S(Y) = S(Y) +
1: IF X > H1% THEN H1% = X
15030 IF Y > H2% THEN H2% = Y
15040 NEXT : CH = 0: V = H1% * H2%: IF V = 1 AND C% < = 30 THEN
GOSUB 770: GOTO 15080
15050 IF V = 1 AND C% > 30 THEN CH = C% * ( ABS (N%(0,0) *
N%(1,1) - N%(0,1) * N%(1,0)) - C% / 2) ^ 2 / T%(0) / S(0) /
S(1): GOTO 15070
15060 FOR I = 0 TO H1%: FOR J = 0 TO H2%: CH = CH + (N%(I,J)
- S(J) * T%(I) / C%) ^ 2 / S(J) / T%(I) * C%: NEXT : NEXT
15070 GOSUB 660: GOSUB 630: PRINT : PRINT : PRINT "FREIHEITSGRADE":;
HTAB 25: PRINT H1% * H2%: PRINT : PRINT "TESTWERT CHI":; HTAB
25: PRINT FN R(CH): PRINT : PRINT "SIGNIFIKANZ":; HTAB 25:
PRINT S
15080 GOSUB 650: GOSUB 700: GOTO 1000

```

```

15000 A# = "KOLMOGOROFF/SMIRNOW-2 STICHPROBEN-TEST": GOSUB 600: VCK%
= 2: PRINT "HINWEIS: VARIABLE 2 MUSS 0/1 CODIERT SEIN.": GOSUB
700: HOME : GOSUB 420: GOSUB 520: IF C% = 0 THEN 710
16010 GOSUB 720: GOSUB 670: DUX = C% / 4: GOSUB 610: N = C%:
GOSUB 100: GOSUB 660: FOR J = 1 TO C%: Y = Y(J): TX(J) = TX(J)
+ 1: NEXT J: H8% = TX(0): H9% = TX(1): H1% = 0: N = H8%: IF H8%
H9% THEN H1% = 1: N = H9%
16020 Q = C% - N: H2% = ABS (H1% - 1): H1 = X * C% + 1: FOR J
= 1 TO 1 STEP - 1: X = X(J): Y = Y(J): IF X = H1 THEN X(J) =
X(J) + 1: Y(J) = Y(J) + 1: GOTO 16040
16030 Y(J) = TX(H1Q) / N: X(J) = TX(H2Q) / Q
16040 TX(Y) = TX(Y) - 1: H1 = X: NEXT J: H1 = Q: H1 = H8%
* H9% / C%: GOSUB 660: GOSUB 630: PRINT : PRINT "CODE 0: "; H8%:;
HTAB 18: PRINT "CODE 1: "; H9%: PRINT : GOSUB 960: GOSUB 950:
GOSUB 700: GOTO 1000
15999 REM

```

```

17000 A# = "FISHER-TEST": GOSUB 600: PRINT "HINWEIS: VARIABLEN
MUESSEN 0/1 CODIERT SEIN": VCK% = 2: GOSUB 700: HOME : GOSUB 420:
GOSUB 520: DUX = C%: GOSUB 610: IF C% = 0 THEN 710
17010 GOSUB 720: FOR J = 1 TO C%: A(0,J) = Y(J): A(1,J) = 0:
NEXT J: GOSUB 720: FOR J = 1 TO C%: A(0,J) = A(0,J): NEXT J: GOSUB
660: GOSUB 670: IF C% < 31 THEN 17020
17015 CALL - 190: HOME : HTAB 8: HTAB 4: PRINT "N IST 20 GRUEN,
ES WIRD EIN": PRINT : HTAB 4: PRINT "CHI QU.-HOMOGENITAETSTEST":
PRINT : HTAB 4: PRINT "DURCHGEFUEHRT.": FOR J = 1 TO 2000: NEXT J
: POKE 34,2: HOME : A# = "CHI QU.-HOMOGENITAETSTEST": GOSUB 600:
GOTO 15010
17020 H3% = 1: GOSUB 830: GOSUB 630: PRINT : HTAB 9: PRINT "TESTWERT 1: ";
: HTAB 25: PRINT "0' 1'": PRINT : PRINT V#(VH%(2)):; GOSUB
840: GOSUB 770: GOSUB 650: GOSUB 700: GOTO 1000
17999 REM

```

```

18000 A# = "MANN/WHITNEY-U-TEST": GOSUB 600: PRINT "HINWEIS: VARIABLE 2
MUSS 0/1 CODIERT SEIN": VCK% = 2: GOSUB 700: HOME : GOSUB 420:
GOSUB 520: HOME : IF C% = 0 THEN 710
18010 GOSUB 720: DUX = C% / 2.5: GOSUB 610
18020 N = C%: GOSUB 660: GOSUB 670: GOSUB 100: GOSUB 180: FOR
J = 1 TO C%: H8% = Y(J): S(H8%) = S(H8%) + X(J): TX(H8%) = TX(H8%)
+ 1: NEXT J: H7 = TX(0) * TX(1): H1 = H7 + TX(0) * (TX(0) + 1)
/ 2 - S(0): IF H1 > H7 / 2 THEN H1 = H7 - H1
18030 Z = (H1 - H7 / 2) / SQR (H7 / C% / (C% - 1) * (C% * C%
* C% - C% - H3) / 12): S = FN R(2 * FN STN(Z)): GOSUB 630:
PRINT : H1% = 1: GOSUB 750
18040 PRINT : PRINT "U "; H1:; HTAB 15: PRINT "ANZ. DER TIES
"; H4%: PRINT : PRINT "TESTWERT 2":; HTAB 25: PRINT FN S(2):;
PRINT : PRINT "SIGNIFIKANZ":; HTAB 25: PRINT S: PRINT : GOSUB
650: GOSUB 700: GOTO 1000

```

```

19000 A# = "KRUSKAL/WALLIS-H-TEST": GOSUB 600: PRINT "HINWEIS:
VARIABLE2 MUSS GANZZAHLIG,": PRINT "VON 0 BIS MAXIMAL 5 CODIERT
SEIN.": VCX = 2: GOSUB 700: HOME: GOSUB 420: GOSUB 520: HOME
: IF CX = 0 THEN 710
19010 GOSUB 670: GOSUB 660: DUX = CX / 3.5: GOSUB 110: N = C1:
GOSUB 100: GOSUB 180: FOR J = 1 TO CX: X = X(0, J) + 1: Y = Y(0, J)
= S(Y) + X: TX(Y) = TX(Y) + 1: IF Y > H1% THEN H1% = Y
19020 NEXT J: H9 = H3: CH = 0: FOR I = 0 TO H1%: IF S(I) > 0 THEN
H2% = H2% + 1: CH = CH + S(I) * S(I) / TX(I)
19030 NEXT I: H2% = H2% - 1: CH = CH * 12 / N / (N + 1) - 3 *
(N + 1): CH = CH / (1 - H9 / (N + N * N - 10)) - H2%: GOSUB
260
19040 GOSUB 630: PRINT: GOSUB 750: PRINT "FREIHEITSGRADE":;
HTAB 25: PRINT H2%:; HTAB 30: PRINT "TIES":; H4%: PRINT: PRINT
"TESTWERT CHI":; HTAB 25: PRINT FN R(CH): PRINT
19050 PRINT "SIGNIFIKANZ":; HTAB 25: PRINT 9: GOSUB 650: GOSUB
700: GOTO 1000
19999 REM

```

```

20000 A# = "MCNEMAR": GOSUB 600: VCX = 2: GOSUB 420: DUX = CX
/ 7.7: GOSUB 520: GOSUB 670: GOSUB 660: GOSUB 720: IF CX = 0
THEN 710
20010 FOR J = 1 TO CX: A#(J) = 1: J1 = 1: J2 = 1: NEXT J: GOSUB
720: GOSUB 610: FOR J = 1 TO CX: Y = A#(J): J1 = 1: J2 = 1: N1(X, Y)
= NX(X, Y) + 1: NEXT J: H3% = NX(0, 1) + NX(1, 0): IF H3% < 10 THEN
CH = (ABS (NX(0, 1) - NX(1, 0)) - 1) ^ 2 / H2%: H1% = 1: GOSUB 260:
GOTO 20040
20020 K = NX(0, 1): IF NX(1, 0) < K THEN K = NX(1, 0)
20030 H9 = H3%: F = .5: GOSUB 290: S = FN R(-F * S)
20040 GOSUB 630: PRINT: HTAB 17: PRINT V$(VCX(UMX(1))): HTAB
13: PRINT "0":; HTAB 21: PRINT "1": GOSUB 20100: HTAB 7: PRINT
"0":; HTAB 9: PRINT "1":; HTAB 12: PRINT NX(0, 0):; HTAB 17:
PRINT "1":; HTAB 20: PRINT NX(1, 0):; HTAB 25: PRINT "1": PRINT
V$(VCX(UMX(2))):; GOSUB 20100
20050 HTAB 7: PRINT "1":; HTAB 9: PRINT "1":; HTAB 12: PRINT
NX(0, 1):; HTAB 17: PRINT "1":; HTAB 20: PRINT NX(1, 1):; HTAB
25: PRINT "1": GOSUB 20100: PRINT: PRINT: IF H3% > 10 THEN
PRINT "TESTWERT CH":; HTAB 25: PRINT FN R(CH): PRINT
20060 PRINT "SIGNIFIKANZ ":; HTAB 25: PRINT FN R(S): PRINT
: GOSUB 650: GOSUB 700: GOTO 1000
20100 HTAB 9: PRINT "I-----I-----I": RETURN

```



```

21000 A# = "COCHRAN-Q-TEST": GOSUB 600:VCX = SX: IF VCX > 5
THEN CALL - 199: VTAB 10: PRINT "ZUVIELE GRUPPEN, ABBRUCH":
GOSUB 700: GOTO 1000
21010 GOSUB 420: GOSUB 560: GOSUB 670:DUX = CX / 2.5: FOR J =
1 TO CX:X(J) = 0: NEXT : GOSUB 660: IF CX = 0 THEN 710
21020 GOSUB 610: FOR J = 1 TO CX: FOR I = 1 TO VCX:X = X(J) + 1:
IF X < 0 AND X < 1 THEN HOME : VTAB 10: PRINT "UNRUNDHEIT
NICHT 0/1 CODIERT. ABBRUCH": GOSUB 700: GOTO 1000
21030 TX(I) = TX(I) + X:X(J) = X(J) + X: NEXT :H1% = H1% + X(J):H2%
= H2% + X(J) * X(J): NEXT : FOR I = 1 TO VCX:H3% = H3% + TX(I):H4%
= H4% + TX(I) * TX(I): NEXT :U = VCX - 1:H1 = VCX * H1% / U:
IF H1 < = 0 THEN S = 1CX:CH = S: GOTO 21050
21040 CH = U * (VCX * H4% - H3% * H2%) / H1: GOSUB 260
21050 GOSUB 640: POKE 33,20: POKE 32,20: HOME : PRINT "CODE
0":; HTAB 10: PRINT "CODE 1": PRINT : FOR I = 1 TO VCX: HTAB
2: PRINT CX - TX(I):; HTAB 11: PRINT TX(I): NEXT : PRINT "PRINT
: POKE 33,40: POKE 32,0: PRINT : PRINT : PRINT "FREIHEITSGRAD":
21060 HTAB 25: PRINT VCX - 1: PRINT : PRINT "TESTWERT 0":;
HTAB 25: PRINT FN R(CH): PRINT : PRINT "SIGNIFIKANZ":; HTAB
25: PRINT S: GOSUB 650:NCX = 0: FOR I = 1 TO 5:VCX(I) = 0:
: GOSUB 700: GOTO 1000
21999 REM

```

```

22000 A# = "WILCOXON-TEST FUER VERSUNDENE PAARE": GOSUB 600: CX
= 2: GOSUB 420: GOSUB 520: HOME : IF CX = 0 THEN 710
22010 GOSUB 660: GOSUB 670:DUX = CX / 4: GOSUB 610: FOR J =
1 TO CX:H1 = X(J) - Y(J): IF SGN(H1) = 0 THEN 22030
22020 H9% = H9% + 1:X:H9% = ABS(X1):Y:H9% = SGN(H1) * Y:
Y:H9% = 1 THEN H1% = H1% + 1
22030 NEXT :N = H9%: GOSUB 100: GOSUB 130: FOR J = 1 TO H9 :S(1)
= S(1) + X(J) * SGN(Y(J) + 1):S(0) = S(0) - X(J) * SGN(Y(J)
- 1): NEXT : IF H1% = 0 OR H1% = H9% THEN S = 1CX:Z = S: GOTO
22050
22040 H2% = H9% - H1%:H3 = S(0): IF S(1) < S(0) THEN H3 = S(1)
21050 Z = (H3 - N * (N + 1) / 4) / SQRT(N * (N + 1) * (2 *
N + 1) / 24):S = FN R(Z) * FN STN(Z)
22060 GOSUB 630: PRINT : PRINT : PRINT "ANZ. DER TIES":; H1%
25: PRINT H4%: PRINT : PRINT "+ DIFFERENZEN ":H1%: HTAB 19:
PRINT "MITTL. RANG ": FN R(S(1) / H1%): PRINT
22070 PRINT "- DIFFERENZEN ":H2%:; HTAB 19: PRINT "MITTL. RANG
": FN R(S(0) / H2%): PRINT : PRINT "TESTWERT Z":; HTAB 25: PRINT
FN R(Z): PRINT : PRINT "SIGNIFIKANZ":; HTAB 25: PRINT S: PRINT
: GOSUB 650: GOSUB 700: GOTO 1000

```

```

25000 A* = "FRIEDMAN": GOSUB 600:VCX = S%: IF VCX > 5 THEN
CALL - 190: VTAB 10: PRINT "ZURUECKE GRUPPEN. ABBRUCH": GOSUB
700: GOTO 1000
23010 GOSUB 420: GOSUB 660: GOSUB 670:P = 0:DUX = CX / 2: GOSUB
610: GOSUB 540: IF CX = 0 THEN 710
23020 FOR K = 1 TO CX: FOR I = 1 TO VCX: YI = X(CX+1,I): YI
= I: NEXT I:N = VCX: GOSUB 100: GOSUB 180: FOR I = 1 TO CX:VCX
= Y(I):S*(H1X) = S*(H1X) + X(I): NEXT
23030 NEXT K: FOR I = 1 TO VCX:P = P + S(I) * S(I): NEXT I:CH
= 12 / CX / VCX / (VCX + 1) * P - 3 * CX * (VCX + 1):V = VCX
- 1: GOSUB 260
23040 GOSUB 640: POKE 33,20: POKE 32,20: HOME : PRINT "MITTLERE
RAENGE": PRINT : FOR I = 1 TO VCX: PRINT FN R*(S(I) / CX): NEXT
: POKE 32,0: POKE 33,40: PRINT : PRINT : PRINT "FREIHEITSGRADE
": HTAB 25: PRINT VCX - 1: PRINT
23050 PRINT "TESTWERT CHI ";; HTAB 25: PRINT FN R*(CH): PRINT
: PRINT "SIGNIFIKANZ ";; HTAB 25: PRINT S: GOSUB 500:NCK = 0:
FOR I = 1 TO 5:VCX(I) = 0: NEXT : GOSUB 700: GOTO 1000
23999 REM

24000 A* = "BRAVAIS/PEARSON": GOSUB 600:VCX = 1: GOSUB 420:
GOSUB 520: HOME : IF CX = 0 THEN 710
24010 DUX = CX / 8: GOSUB 610: GOSUB 670: FOR J = 1 TO CX:
= X(J):Y = Y(J):H2 = H2 + X:H3 = H3 + 1:H4 = H4 + X * X:H5 =
H5 + Y * Y:H1 = H1 + X * Y: NEXT J:H5 = (H4 - H2 * H2 / CX) *
(H5 - H3 * H3 / CX): IF H5 = 0 THEN 740
24020 Z = (H1 - H2 * H3 / CX) / SQR (H5 - 2 * H3 * H3 / CX): GOSUB 210: GOSUB 530: PRINT
PRINT : PRINT "KORRELATIONSDEFF. F";; HTAB 25: PRINT FN R*(Z)
PRINT
24030 PRINT "FREIHEITSGRADE";; HTAB 25: PRINT V: PRINT : PRINT
"TESTWERT T";; HTAB 25: PRINT FN R*(T): PRINT : PRINT "SIGNIFIKANZ";;
HTAB 25: PRINT S: PRINT : GOSUB 650: GOSUB 700: GOTO 1000

```



```

25000 A$ = "SPEARMAN": GOSUB 600:VC% = 2: GOSUB 420: GOSUB 520:
HOME : IF CX = 0 THEN 710
25010 GOSUB 670:DU% = CX / 2: GOSUB 610:N = CX: GOSUB 100:
GOSUB 180:H9 = H3:H9% = H4%: FOR J = 1 TO CX:A(0,J) = X(J):X(J)
= Y(J):Y(J) = A(0,J): NEXT : GOSUB 100: GOSUB 190: FOR J = 1
TO CX:A(0,J) = X(J):X(J) = Y(J):Y(J) = A(0,J): NEXT
25020 FOR J = 1 TO CX:H8 = H8 + (X(J) - Y(J))^ 2: NEXT :H1
= N * N * N:H9 = (H1 - N - H9) / 12:H3 = (H1 - N - H9) / 12:
IF H9 * H3 = 0 THEN 2 = 1C%:S = 2: GOTO 25040
25030 Z = (H9 + H3 - H8) / 2 / SQR (H9 * H3):U = N - 2:T =
2 * SQR (ABS (U / (1 - Z * Z))): GOSUB 210
25040 GOSUB 630: PRINT : PRINT : HTAB 15: PRINT "VARIABLE1
VARIABLE2": PRINT : PRINT "ANZ. DER TIES": HTAB 18: PRINT
H9%: HTAB 29: PRINT H4%: PRINT : PRINT : PRINT "KORRELATIONSKOEFF.":
HTAB 25: PRINT FN R(2): PRINT
25050 PRINT "FREIHEITSGRADE": HTAB 25: PRINT N - 2: PRINT
: PRINT "TESTWERT T": HTAB 25: PRINT FN R(1): PRINT : PRINT
"SIGNIFIKANZ": HTAB 25: PRINT S: PRINT : GOSUB 650: GOSUB 700:
GOTO 1000
25999 REM

```

```

26000 A$ = "CHI QU.-UNABHAENGIGKEITSTEST": GOSUB 600: PRINT
"HINWEIS: DIE VARIABLEN SOLLEN": PRINT "GANZZAHLIG, VON 0 BIS
HOECHSTENS 5,": PRINT "CODIERT SEIN.":VC% = 2: GOSUB 700: HOME
: GOSUB 420: GOSUB 520: HOME : IF CX = 0 THEN 710
26010 GOSUB 670: GOSUB 660:CH = 0:DU% = CX / 4: GOSUB 610:
FOR J = 1 TO CX:X = X(J):Y = Y(J): IF X > H1 THEN H1 = X
26020 IF Y > H2 THEN H2 = Y
26030 IF H1 > 5 OR H2 > 5 THEN HOME : UTAB 10: PRINT "FALSCHES
CODIERUNG. ABRUCH": GOSUB 700: GOTO 1000
26040 NX(X,Y) = NX(X,Y) + 1: NEXT : FOR I = 0 TO H1: FOR J =
0 TO H2:S(I) = S(I) + NX(I,J): NEXT : NEXT : FOR J = 0 TO H2:
FOR I = 0 TO H1:T%(J) = T%(J) + NX(I,J): NEXT : NEXT
26050 FOR J = 0 TO H2: FOR I = 0 TO H1:H3 = S(I) * T%(J)
CX:CH = CH + (NX(I,J) - H3) ^ 2 / H3: NEXT : NEXT :U = H3 *
H2:H1% = H1:H2% = H2: GOSUB 260
26060 GOSUB 630: PRINT : PRINT : PRINT "FREIHEITSGRADE": HTAB
25: PRINT H1% * H2%: PRINT : PRINT "TESTWERT CH": HTAB 25:
PRINT FN R(CH): PRINT : PRINT "SIGNIFIKANZ": HTAB 25: PRINT
S: PRINT
26070 INPUT "HAEUFIGKEITEN GEWUNSCHT ? (J/N) ":A$: IF A$ =
> "J" AND A$ < > "N" THEN CALL - 998: CALL - 958: GOTO
26070
26080 IF A$ = "N" THEN CALL - 998: CALL - 958: GOSUB 650:
GOSUB 700: GOTO 1000
26090 HOME : HTAB 12: PRINT V$(VC%(UM%(2))): PRINT : HTAB 4:
PRINT "CODE ": FOR J = 0 TO H2%: HTAB (10 + 5 * J): PRINT
J: NEXT : PRINT : PRINT : PRINT V$(VC%(UM%(1))): PRINT : FOR
I = 0 TO H1%: PRINT "CODE ":I:
26100 FOR J = 0 TO H2%: HTAB (10 + 5 * J): PRINT NX(I,J):
NEXT : PRINT : NEXT : PRINT : GOSUB 650: GOSUB 700: GOTO 26060

```

```
61000 IF PEEK (222) = 254 THEN HTAB 5: CALL - 198: INVERSE
: PRINT "FALSCH EINGABE. WIEDERHOLUNG": FOR Q = 1 TO 1500:
NEXT : NORMAL : CALL - 998: CALL - 998: CALL - 958: RESUME

61010 HOME : POKE 768,104: POKE 769,168: POKE 770,100: POKE
771,166: POKE 772,228: POKE 773,154: POKE 774,72: POKE 775,152:
POKE 776,72: POKE 777,96: CALL 768

61020 HOME : IF PEEK (222) = 5 THEN VTAB 10: CALL - 198:
INVERSE : HTAB 12: PRINT "FALSCH DISKETTE": NORMAL : GOTO 61060

61030 IF PEEK (222) = 6 THEN VTAB 10: CALL - 198: HTAB 12:
PRINT "FALSCH DISKETTE": GOTO 61060

61040 IF PEEK (222) = 8 THEN VTAB 10: HTAB 4: CALL - 198:
PRINT "LESEN/SCHREIBEN UNMOEGLICH. WIEDERHOLUNG": GOTO 61060

61050 VTAB 9: CALL - 958: VTAB 10: HTAB 12: PRINT "FATAL ERROR":
PRINT : HTAB 7: PRINT "NEUSTART DES SYSTEMS": VTAB 22: END

61060 GOSUB 700: CLEAR : GOTO 10
```

Definition einer Datei : FILE INPUT


```

*****
*                                     *
*               ***REP-TEST***       *
*                                     *
*   NICHT-PARAMETRISCHE TESTVERFAHREN *
*                                     *
*           (C) RENATE PRUST 1983     *
*                                     *
*           FILE INPUT                *
*                                     *
*****

```

```

10 HOME
20 REM *****
30 REM *
40 REM * DATAN II FILE INPUT *
50 REM *
60 REM * RAINER SCHNELL 3/82 *
70 REM * MODIFIED 1/9/82 *
80 REM *****
90 REM

```

```

100 TEXT : HOME : SPEED= 250: NOTRACE :D$ = CHR$ (4): PRINT
D$"NOMONI"

```

```

110 DIM V$(80):IS% = 1:IN% = 1
120 REM

```

FILE NAME INPUT

```

130 HOME : GOSUB 690: HTAB (1): VTAB (6): PRINT "FILE NAME
(MAX.6 CHARACTERS)": INPUT F$: IF LEN (F$) > 6 THEN PRINT
CHR$ (7): GOTO 130
140 IF F$ = "" OR F$ = " " THEN PRINT CHR$ (7): GOTO 130
150 IF ASC ( LEFT$ (F$,1)) < 65 OR ASC ( LEFT$ (F$,1)) >
90 THEN PRINT CHR$ (7): INVERSE : PRINT "FIRST CHARACTER MUST
BE A LETTER": NORMAL : FOR W = 1 TO 1000: NEXT W: GOTO 130

```

160 REM

NUMBER OF VARIABLES

```
170 GOSUB 690: HTAB (1): VTAB (6): PRINT "NUMBER OF VARIABLES  
": INPUT NV%: IF NV% < 2 OR NV% > 80 THEN PRINT CHR$(7):  
GOTO 170  
180 REM
```

N OF CASES

```
190 GOSUB 690: HTAB (1): VTAB (6): PRINT "N OF CASES ": INPUT  
NF%: IF NF% < 2 OR NF% > 200 THEN PRINT CHR$(7): GOTO 190  
200 REM
```

DISPAY FILE PARAMETER

```
210 GOSUB 690: HTAB (1): VTAB (6): PRINT "FILE NAME.....":  
HTAB (22): VTAB (6): PRINT F$: VTAB (8): HTAB (1): PRINT "NUMBER  
OF VARIABLES..": VTAB (8): HTAB (22): PRINT NV%  
220 HTAB (1): VTAB (10): PRINT "NUMBER OF CASES.....": HTAB  
(22): VTAB (10): PRINT NF%  
230 VTAB (15): HTAB (22): PRINT "OKAY ": INPUT A$: IF LEFT$(  
A$,1) = "N" THEN GOTO 130  
240 REM
```

VARIABLE NAME INPUT

```
250 FOR U = 1 TO NV%  
260 GOSUB 690: HTAB (1): VTAB (6): PRINT "NAME OF VARIABLE  
NO.": U: HTAB (25): VTAB (6): INPUT A$  
270 IF LEN (A$) > 6 THEN HTAB (25): VTAB (6): INVERSE : PRINT  
"MAX.6 CHARACTERS": GOSUB 700: GOTO 260  
280 IF ASC (A$) < 65 OR ASC (A$) > 90 THEN HTAB (18): VTAB  
(6): INVERSE : PRINT "FIRST MUST BE A LETTER": GOSUB 700: GOTO  
260  
290 INVERSE : HTAB (25): VTAB (6): PRINT "CHECK NAME": NORMAL  
  
300 FOR H = 1 TO NV%: IF V$(H) = A$ AND H < > U THEN INVERSE  
: HTAB (25): VTAB (6): PRINT "ALREADY EXISTS": GOSUB 700: GOTO  
260  
310 NEXT H: V$(U) = A$: NEXT U
```

320 REM

LIST VARS

```

330 GOSUB 690: PRINT :PC = 0: FOR U = 1 TO NV%:PC = PC + 1
340 IF PC = 1 THEN GOSUB 690: HTAB (1): VTAB (6): PRINT "VARIABLE
NO.":; HTAB (20): PRINT "NAME": PRINT
350 HTAB (5): PRINT U:; HTAB (9): PRINT ".....":; HTAB
(23): PRINT V$(U)
360 IF PC = 6 THEN PC = 0: VTAB (20): HTAB (20): PRINT "OKAY":;
INPUT A$: IF LEFT$(A$,1) = "N" THEN GOTO 410
370 PRINT : NEXT U
380 VTAB (20): HTAB (20): PRINT "OKAY":; INPUT A$: IF LEFT$(
A$,1) = "N" THEN GOTO 410
390 GOTO 490
400 REM

```

LABEL CORRECTION

```

410 HTAB (5): VTAB (20): PRINT "EDIT WHICH LABEL (NO.):": INPUT
CX: IF CX < 1 OR CX > NV% THEN PRINT CHR$(7): GOTO 410
420 GOSUB 690: HTAB (1): VTAB (4): PRINT "OLD LABEL":; HTAB
(20): VTAB (4): PRINT "NEW LABEL"
430 HTAB (1): VTAB (6): PRINT V$(CX):; HTAB (20): INPUT A$
440 IF LEN(A$) > 6 THEN PRINT : INVERSE : PRINT "MAX.6 CHARACTERS"
: GOSUB 700: GOTO 420
450 IF ASC(A$) < 65 OR ASC(A$) > 90 THEN PRINT CHR$(7):
INVERSE : PRINT "FIRST MUST BE A LETTER": GOSUB 700: GOTO 420
460 FOR H = 1 TO NV%: IF V$(H) = A$ THEN PRINT : INVERSE :
PRINT "ALREADY EXISTS": GOSUB 700: GOTO 420
470 NEXT H:V$(CX) = A$: GOTO 330
480 REM

```

DEFINE MISSING VALUE

```

490 GOSUB 690: VTAB (6): HTAB (10): PRINT "MISSING VALUE CODE"
500 PRINT : PRINT "THE CODE IS VALIDE FOR ALL VARIABLES"
510 PRINT : PRINT "ENTER NUMERICAL CODE OF MISSING VALUES"
520 VTAB 15: PRINT "MISSING CODE ":; HTAB 25: INVERSE : PRINT
"DEFAULT=999.99":; NORMAL : HTAB 14: INPUT A$: IF A$ = "" THEN
MW = 999.99: GOTO 590
530 MW = VAL(A$)
590 ONERR GOTO 710
600 GOSUB 690: PRINT CHR$(7): VTAB (6): HTAB (10): PRINT
"INSERT NEW DATA DISK": PRINT : PRINT "THE DATA DISK MUST BE
BLANK BUT IT MUST"
610 PRINT : PRINT "BE INITIALIZED.": PRINT : INPUT "PRESS RETURN
AFTER INSERTION":A$

```

620 REM

WRITING VARCATALOG

```
630 GOSUB 690: VTAB (6): HTAB (6): PRINT "WRITING VARIABLECATALOG":  
VTAB (10): HTAB (6): PRINT "DURATION: MAX.10 SECONDS"  
640 D$ = CHR$(4): PRINT D$"OPEN VARCATALOG": PRINT D$"WRITE  
VARCATALOG": PRINT F$: PRINT NV%: PRINT NF%: PRINT MW: PRINT  
IS%: PRINT IN%: FOR U = 1 TO NV%: PRINT V$(U): NEXT U: PRINT  
D$"CLOSE VARCATALOG"  
650 REM
```

END

```
660 GOSUB 690: VTAB (6): HTAB (16): PRINT "DONE": VTAB (10):  
HTAB (1): PRINT "INSERT DATAN SYSTEM DISK.":: INPUT "PRESS RETURN":A$  
670 HOME: VTAB (10): HTAB (10): PRINT "ZURUECK ZUM MENU":D$  
= CHR$(4): PRINT D$"RUNREP-TEST MENU"  
680 REM
```

COMMON SUBROUTINES

```
690 HOME: PRINT "          < DATAN II FILE INPUT >": RETURN  
700 NORMAL: PRINT CHR$(7): FOR W = 1 TO 1000: NEXT W: RETURN  
  
710 GOSUB 690: VTAB (10): HTAB (12): PRINT "FATAL ERROR": VTAB  
(12): HTAB (10): PRINT "RESTART SYSTEM": VTAB (20): END
```


Dateneingabe : DATA INPUT


```

*****
*
*          ***REP-TEST***
*
*  NICHT-PARAMETRISCHE TESTVERFAHREN
*
*          (C) RENATE PRUST 1983
*
*          DATA INPUT
*
*****

```

```

10 REM *****
20 REM *
30 REM * DATAN II DATA INPUT *
40 REM *
50 REM * RAINER SCHNELL 3/82 *
60 REM *
70 REM *****
80 REM

```

INIT

```

90 DEF FN MOD(A) = INT ((A / 10 - INT (A / 10)) * 10 + 0.05)
100 DEF FN SP(X) = ( FN MOD(X) + (1 - SGN ( FN MOD(X))) ) *
10) * 2 + 2: ONERR GOTO 930
110 TEXT : HOME : SPEED= 250: NOTRACE : D$ = CHR$ (4): PRINT
D$"NO MONI": POKE 34,0: DIM ID(201): DIM A(201): DIM V$(84): VT
= 2: HT = 23: PC = 1: GOSUB 700
120 REM

```

INPUT VARCATALOG AND

CHECK

INPUT STATUS

```

130 GOSUB 610: IF IS% = NV% + 1 THEN GOSUB 650: VTAB (10):
HTAB (10): PRINT "VARIABEL DATA COMPLETE": VTAB (12): HTAB (8):
PRINT "ADDITIONAL INPUT IMPOSSIBLE": VTAB (14): HTAB (8): INPUT
"RETURN TO MENU.PRESS RETURN": A$: GOTO 720

```

140 REM

START INPUT

```

150 FOR U = 1% TO NV%:VT = 2: GOSUB 690: IF IN% < > 1 THEN
  VTAB (10): HTAB (10): PRINT "READING DATA": VTAB (12): HTAB
(7): PRINT "DURATION:6 SECONDS": GOSUB 830: GOSUB 690:PC =
INT (IN% / 10) + 1
160 VTAB (10): PRINT "RANGE CHECK DESIRED": INPUT R$: IF R$
= CHR$ (89) THEN VTAB (10): CALL - 958: VTAB (10): PRINT
"UPPER LIMIT OF VALID CASES": INPUT UL: VTAB (12): PRINT "LOWER
LIMIT OF VALID CASES": INPUT LL
170 IF R$ = CHR$ (89) AND (UL < = LL) THEN PRINT CHR$ (7):
VTAB (8): CALL - 958: GOTO 160
180 IF R$ = CHR$ (89) THEN VTAB (10): CALL - 958: HTAB (10):
PRINT "UPPER LIMIT ":UL: VTAB (12): HTAB (10): PRINT "LOWER
LIMIT ":LL: VTAB (22): HTAB (28): PRINT "OKAY": INPUT A$: IF
A$ < > CHR$ (89) THEN VTAB (10): CALL - 958: GOTO 160
190 VTAB (10): CALL - 958
200 FOR I = IN% TO NF%:VT = FN SP(I)
210 REM

```

INPUT

```

220 VTAB (VT): HTAB (1): PRINT "CASE NR ":I: HTAB (13): PRINT
"....": HTAB (17): INPUT " ":A$
230 REM

```

CHECK INPUT

```

240 IF PEEK (37) > VT THEN VTAB ( PEEK (37)): CALL - 868:
GOSUB 580: GOTO 220
250 IF A$ = "M" THEN A(I) = MW: VTAB (VT): HTAB (18): PRINT
MW: GOTO 290
260 IF STR$ ( VAL (A$)) < > A$ THEN GOSUB 580: GOTO 220
270 IF R$ = CHR$ (89) AND ( VAL (A$) < LL OR VAL (A$) > UL)
THEN GOSUB 590: GOTO 220
280 A(I) = VAL (A$)
290 IO(I) = A(I): IF VT = 22 OR I = NF% THEN GOSUB 340:VT =
2: GOSUB 690:PC = PC + 1
300 PRINT : NEXT I: GOSUB 780: GOSUB 900
310 IF U < NV% THEN GOSUB 660: VTAB (10): HTAB (10): PRINT
"CONTINUE INPUT": INPUT A$: IF A$ = "N" THEN GOTO 750
320 PC = 1: NEXT U:U = U - 1: GOSUB 900: GOTO 730
330 REM

```

CONTROL INPUT DATA

```

340 HTAB (29): VTAB (2): INVERSE : PRINT "CONTROL DATA": NORMAL
: HTAB (30): VTAB (23): PRINT "OKAY": INPUT A$: IF A$ = "Y"
THEN RETURN

```

350 REM

EDIT INPUT DATA

```

360 VTAB (23): HTAB (30): CALL - 868: HTAB (29): VTAB (2):
INVERSE : PRINT "EDIT DATA " : NORMAL
370 VTAB (23): HTAB (30): CALL - 868: VTAB (4): HTAB (29):
CALL - 868: PRINT "EDIT CASE": VTAB (6): HTAB (29): CALL -
868: INPUT "CASE NR. ":EX
380 IF EX < 1 OR EX > NF% THEN PRINT CHR$(7): GOTO 370
390 IF EX < (PC - 1) * 10 + 1 OR EX > (PC * 10) THEN GOSUB
560: GOTO 370
400 IF IN%( > 1 AND (EX < IN%) THEN GOSUB 560: GOTO 370
410 VTAB (4): HTAB (29): CALL - 868: VTAB (6): HTAB (29):
CALL - 868
420 REM

```

SET CASE TO EDIT

```

430 HI = 1: I = EX: VT = FN SP(EX)
440 VTAB (VT - 1): HTAB (29): INVERSE : PRINT "ENTER DATA":
NORMAL
450 VTAB (VT): HTAB (29): INPUT A$: IF PEEK (30) : VT THEN
VTAB (VT + 1): HTAB (1): CALL - 868: GOTO 520
460 VTAB (VT - 1): HTAB (29): CALL - 868
470 IF A$ = "M" THEN A(I) = MW: GOTO 510
480 IF R$ = CHR$(89) AND (VAL (A$) < LL OR VAL (A$) > UL)
THEN GOSUB 590: GOTO 440
490 IF STR$( VAL (A$)) < > A$ THEN GOTO 520
500 A(I) = VAL (A$)
510 IO(I) = A(I): HTAB (29): CALL - 868: VTAB (VT): HTAB (18):
CALL - 868: PRINT A(I): VTAB (VT): HTAB (29): CALL - 868: I
= HI: GOTO 340
520 VTAB (VT - 1): HTAB (29): CALL - 868
530 VTAB (VT): HTAB (18): CALL - 868: PRINT CHR$(7): VTAB
(VT): HTAB (18): INVERSE : PRINT "INVALID INPUT": VTAB (VT +
1): HTAB (18): CALL - 868
540 FOR W = 1 TO 500: NEXT W: VTAB (VT): HTAB (18): CALL -
868: GOTO 440
550 REM EDIT ERROR MESSAGE
560 PRINT CHR$(7): VTAB (4): HTAB (29): INVERSE : CALL -
868: PRINT "NOT ON": VTAB (6): HTAB (29): CALL - 868: PRINT
"SCREEN": NORMAL : FOR W = 1 TO 500: NEXT W: RETURN
570 REM

```

INVALID INPUT ROUTINE

```

580 VTAB (VT): HTAB (HT): INVERSE : PRINT " INVALID INPUT":
PRINT CHR$(7): FOR W = 1 TO 500: NEXT W: NORMAL : VTAB (VT):
HTAB (HT): CALL - 958: RETURN
590 VTAB (VT): HTAB (HT): INVERSE : PRINT " RANGE ERROR " :
PRINT CHR$(7): FOR W = 1 TO 500: NEXT W: NORMAL : VTAB (VT):
HTAB (HT): CALL - 868: RETURN

```

600 REM

INPUT VARCATALOG

```
610 GOSUB 650: VTAB (10): HTAB (10): PRINT "READING VARCATALOG":
VTAB (12): HTAB (6): PRINT "DURATION: MAX.15 SECONDS"
620 D$ = CHR$ (4): PRINT D$"OPEN VARCATALOG": PRINT D$"READ VARCATALOG":
INPUT F$: INPUT NV%: INPUT NF%: INPUT MW: INPUT IS%: INPUT
IN%: FOR U = 1 TO NV%: INPUT V$(U): NEXT U: PRINT D$"CLOSE VARCATALOG":
RETURN
630 REM
```

COMMON SUBROUTINES

```
640 HOME : RETURN
650 HOME : HTAB (8): PRINT "<DATAN II DATA INPUT>": RETURN
660 HOME : PRINT " < DATAN II DATA INPUT FILE ";F$;" > ": RETURN
670 HOME : PRINT "VARIABLE ";V$(U): PRINT : RETURN
680 GOSUB 640: VTAB (2): HTAB (2): PRINT "VARIABLE...";V$(U)::
HTAB (21): PRINT "MODE...": RETURN
690 GOSUB 680: HTAB (29): VTAB (2): INVERSE : PRINT "INPUT
DATA": NORMAL : RETURN
700 GOSUB 650: PRINT CHR$ (7): VTAB (10): HTAB (3): INVERSE
: PRINT "INSERT DATA DISK AND PRESS RETURN": NORMAL : INPUT
":A$: RETURN
710 GOSUB 640: PRINT CHR$ (7): VTAB (10): HTAB (3): INVERSE
: PRINT "INSERT SYSTEM DISK AND PRESS RETURN": NORMAL : INPUT
":A$: RETURN
720 REM
```

END

```
730 GOSUB 660: VTAB (6): HTAB (10): PRINT "DATA INPUT COMPLETED"
740 PRINT : PRINT "ALL VARIABLE FILES HAD BEEN CREATED": FOR
W = 1 TO 2000: NEXT W
750 GOSUB 710: HOME : VTAB (10): HTAB (10): PRINT "ZURUECK
ZUM MENU"
760 PRINT CHR$ (4):"RUN REP-TEST MENU"
770 REM
```

WRITING DATA

```
780 N$ = V$(U) + " VARDATA FILE " + F$
790 GOSUB 860
800 GOSUB 660: VTAB (10): HTAB (5): PRINT "WRITING ";N$: VTAB
(12): HTAB (10): PRINT "DURATION: 5 SECONDS"
810 D$ = CHR$ (4): PRINT D$"BSAVE";N$;"A";SI;"L1017": RETURN
```

820 REM

READING DATA

```
830 N$ = V$(U) + " VARDATA FILE " + F$: GOSUB 860: GOSUB 840:
FOR I = 1 TO NF$:A(I) = IO(I): NEXT I: RETURN
840 D$ = CHR$(4): PRINT D$"BLOAD";N$;";A":SI: RETURN
850 REM
```

COMPUTING

POSITION IO

```
860 ST% = PEEK (107) + PEEK (108) * 256:EN% = PEEK (109)
+ PEEK (110) * 256
870 FOR SI = ST% TO EN%: IF PEEK (SI) = 73 AND PEEK (SI +
1) = 79 AND PEEK (SI + 2) = 249 AND PEEK (SI + 3) = 3 THEN
RETURN
880 NEXT SI: GOTO 930: REM IO NOT FOUND
890 REM
```

UPDATING STATUS POINTER

```
900 GOSUB 660: VTAB (10): HTAB (8): PRINT "UPDATING STATUS
POINTER": VTAB (12): HTAB (10): PRINT "DURATION:3 SECONDS"
910 D$ = CHR$(4): PRINT D$"OPEN VARCATALOG": PRINT D$"POSITION
VARCATALOG,R5": PRINT D$"WRITEVARCATALOG": PRINT U + 1: PRINT
D$"CLOSE VARCATALOG": RETURN
920 REM
```

ERROR ROUTINE

```
930 GOSUB 660: PRINT CHR$(7): PRINT CHR$(7): VTAB (10):
HTAB (15): PRINT "FATAL ERROR": VTAB (14): HTAB (10): PRINT
"RESTART DATAN SYSTEM": VTAB (18): END
```


Datenverwaltung : DATA BASE MANAGEMENT SYSTEM


```

100 GOSUB 190: GOSUB 180: VTAB (3): HTAB (2): PRINT "DATA INPUT":;
HTAB (22): PRINT "DATA EDITING"
110 VTAB (5): PRINT "FILE INPUT.....1":; HTAB (21): PRINT
"LIST CASES.....3": VTAB (7): PRINT "DATA INPUT.....2":;
HTAB (21): PRINT "EDIT DATA.....4": GOSUB 180
120 VTAB (9): HTAB (2): PRINT "FILE MANAGEMENT":; HTAB (22):
PRINT "DATA MODIFICATION": VTAB (11): PRINT "ADD CASES.....5":;
HTAB (21): PRINT "COMPUTE.....10": VTAB (13): PRINT "ADD
VARIABLES....6":; HTAB (21)
130 PRINT "COUNT.....11": VTAB (15): PRINT "DELETE VARIABLE..7
":; HTAB (21): PRINT "RECODE.....12": VTAB (17): PRINT
"DELETE FILE.....8":; HTAB (21): PRINT "IF.....13"
140 VTAB (19): PRINT "COPY FILE.....9":; HTAB (21): PRINT
"2-TRANSFORMATION.14": VTAB (20): GOSUB 180: VTAB (21): PRINT
"CLEAR MEMORY.....15":; HTAB (21): PRINT "MAIN MENU.....17"
150 VTAB (23): PRINT "FILE CATALOG.....16":; HTAB (21): INPUT
"WHICH OPTION.....":A%: IF A% < 1 OR A% > 17 THEN PRINT CHR$
(7): GOTO 100
160 IF A% > 2 AND A% < > 15 AND A% < > 9 AND A% < > 17 THEN
GOSUB 740
170 ON A% GOSUB 1000,2000,3000,4000,5000,6000,7000,8000,9000,10000,11
000,12000,13000,14000,15000,16000,17000: GOTO 100
180 FOR I = 1 TO 39: PRINT "-":; NEXT : PRINT : RETURN
190 HOME : HTAB (3): PRINT "DATAN II DATA BASE MANAGEMENT SYSTEM":
RETURN
200 POKE - 16368,0: WAIT - 16384,128: POKE - 16368,0: RETURN

210 GOSUB 190: VTAB (10): HTAB (15): PRINT "DONE": GOSUB 220:
RETURN
220 FOR W = 1 TO 750: NEXT W: RETURN
230 VTAB (23): HTAB (28): PRINT "OKAY <Y/N>":; INPUT A$: RETURN

240 INPUT A$: IF A$ < > CHR$ (89) THEN POP
250 RETURN
260 IF A$ = "" THEN POP
270 RETURN
280 GOSUB 720: HOME : VTAB (10): HTAB (10): RETURN
300 FOR J = 1 TO NF%:A(0,J) = 0: NEXT : RETURN
400 IF NV% = 82 THEN GOSUB 190: VTAB (10): HTAB (5): PRINT
"NO DATA MODIFICATION POSSIBLE": VTAB (12): HTAB (5): PRINT
"TOO MUCH VARIABLES ON DISK": GOSUB 220: POP
410 RETURN
500 VTAB (10): HTAB (10): PRINT "DESIRED VARIABLE ":; INPUT
A$: IF A$ = "" THEN POP : RETURN
510 A% = 0:I% = 0: FOR U = 1 TO NV%: IF A$ = V$(U) THEN GOTO
530
520 NEXT U: GOSUB 190: VTAB (10): HTAB (5):; PRINT "VARIABLE
"/A$;: PRINT "/ DON'T EXIST": GOSUB 220: GOSUB 190: GOTO 500
530 FOR U = 1 TO NCPU%: IF A$ = VCPU$(U) THEN I% = U: RETURN

540 NEXT U

```

```

550 IF NCPU% < 15 THEN NCPU% = NCPU% + 1: I% = NCPU%: GOTO 590
560 GOSUB 190: VTAB (10): PRINT "MEMORY FULL.YOU MUST DELETE
A CPU-VAR.": VTAB (12): PRINT "ENTER NUMBER OF CPU-VAR TO DELETE"
570 VTAB (14): PRINT "<PRESS '0' TO DISPLAY CPU CATALOG.": :
INPUT I%: IF I% < 0 OR I% > 15 THEN 560
580 IF I% = 0 THEN GOSUB 600: GOTO 560
590 VCPU$(I%) = A$: GOSUB 910: RETURN
600 HOME : HTAB (10): PRINT "< CPU VARIABLE CATALOG >"
610 PRINT : PC% = 1: FOR U = 1 TO NCPU%: HTAB (PC%): PRINT "<";U;">
": VCPU$(U): PC% = PC% + 12: IF PC% = 37 THEN PC% = 1: PRINT
: PRINT
620 NEXT U: GOSUB 200: RETURN
700 IF LF% = 1 THEN POP : RETURN
710 GOSUB 190: PRINT CHR$(7): VTAB (10): HTAB (3): INVERSE
: PRINT "INSERT DATA DISK AND PRESS RETURN": NORMAL : INPUT
": A$: LF% = 1: RETURN
720 IF LF% = 0 THEN RETURN
730 GOSUB 190: PRINT CHR$(7): VTAB (10): HTAB (3): INVERSE
: PRINT "INSERT SYSTEM DISK AND PRESS RETURN": NORMAL : INPUT
": A$: RETURN
740 GOSUB 700: GOSUB 190: VTAB (10): HTAB (10): PRINT "READING
VARCATALOG": VTAB (12): HTAB (8): PRINT "MAX.DURATION:15 SECONDS"
750 D$ = CHR$(4): PRINT D$"OPEN ": VQ$: PRINT D$"READ": D$:
INPUT F$: INPUT NV%: INPUT NF%: INPUT MW: INPUT IS%: INPUT IN%:
FOR U = 1 TO NV%: INPUT V$(U): NEXT U: PRINT D$"CLOSE ": VQ$: LF%
= 1: RETURN
760 GOSUB 190: VTAB (10): HTAB (10): PRINT "WRITING VARCATALOG":
VTAB (12): HTAB (8): PRINT "MAX.DURATION:10SECONDS"
770 D$ = CHR$(4): PRINT D$"OPEN": VQ$: PRINT D$"WRITE": VQ$:
PRINT F$: PRINT MW: PRINT NF%: PRINT MW: PRINT IS%: PRINT IN%:
FOR U = 1 TO NV%: PRINT V$(U): NEXT U: PRINT D$"CLOSE": VQ$:
RETURN
800 GOSUB 190: VTAB (10): PRINT "NAME OF RESULT-VARIABLE": :
INPUT A$: GOSUB 820: IF M% = 1 THEN 800
810 IS% = IS% + 1: NV% = NV% + 1: V$(NV%) = A$: RETURN
820 M% = 0: GOSUB 190: VTAB (8): HTAB (4): PRINT "CHECK NAME
": A$: IF LEN (A%) > 6 THEN GOSUB 870: PRINT ".MAX.6 CHARACTERS":
GOTO 880
830 IF A% = "" THEN 880
840 IF ASC (A%) < 65 OR ASC (A%) > 90 THEN GOSUB 870: PRINT
".FIRST MUST BE A LETTER": GOTO 880
850 FOR H = 1 TO NV%: IF V$(H) = A% AND H < > U THEN GOSUB
870: PRINT ".ALREADY EXISTS": GOTO 880
860 NEXT H: RETURN
870 VTAB (10): HTAB (4): CALL - 211: RETURN
880 GOSUB 200: M% = 1: RETURN
900 N$ = A% + F2$ + F$: GOSUB 190: VTAB (10): HTAB (5): PRINT
"WRITING ": N$: VTAB (12): HTAB (10): PRINT "DURATION: 5 SECONDS":
PRINT CHR$(4) "BSAVE" N$: "A": FN ME(107): "L": FN ME( FN ME(107)
+ 2): RETURN

```

```

910 GOSUB 190: VTAB (10): HTAB (10): PRINT "LOADING VARIABLE":
VTAB (12): HTAB (8): PRINT "DURATION: 7 SECONDS":N$ = VCPU$(IX)
+ F2$ + F$: PRINT CHR$(4);"BLOAD";N$;"A"; FN ME(107): FOR
J = 1 TO NF$:A(IX,J) = IO(J): NEXT J: RETURN
950 FOR J = 1 TO NF$:IO(J) = A(0,J): NEXT J: GOSUB 900: GOSUB
760: RETURN
1000 GOSUB 280: PRINT "LOADING FILE INPUT":D$ = CHR$(4):
PRINT D$"RUN DATAN II FILE INPUT"
2000 GOSUB 280: PRINT "LOADING DATA INPUT":D$ = CHR$(4):
PRINT D$"RUN DATAN II DATA INPUT"
3000 GOSUB 3050: VTAB (10): PRINT "ENTER N OF VARIABLES TO
LIST";: INPUT N1$: IF N1% < 1 OR N1% > 15 OR N1% > NV% THEN
3000
3010 FOR G = 1 TO N1%: GOSUB 3050: VTAB (8): HTAB (5): PRINT
"VARIABLE ";G;: PRINT " OF VARIABLE LIST": GOSUB 500:VM%(G)
= I%: NEXT G
3020 FOR J = 1 TO NF%: HOME : HTAB (10): PRINT "CONTENT OF
CASE ";J: VTAB (3): PRINT "VARIABLE";: HTAB (16): PRINT "CODE"
3030 PRINT : PRINT : FOR G = 1 TO N1%: PRINT VCPU$(VM%(G));:
HTAB (7): PRINT ".....": HTAB (16): PRINT A(VM%(G),J): NEXT
: VTAB (24): INPUT "<PRESS RETURN TO CONTINUE,'0'FOR MENU";A$:
IF A$ = "0" THEN RETURN
3040 NEXT J: RETURN
3050 HOME : HTAB (8): PRINT "DATAN II LIST CASES": RETURN
4000 M% = 0: GOSUB 4110: GOSUB 500: GOSUB 4110: VTAB (8): HTAB
(10): PRINT "EDIT OPTIONS": VTAB (10): HTAB (1): PRINT "CASE
IDENTIFICATION BY SEQ.NUM.....1": VTAB (12): HTAB (1): PRINT
"CASE IDENTIFICATION BY RANGE CHECK....2"
4010 VTAB (14): HTAB (20): PRINT "WHICH OPTION";: INPUT A%:
ON A% GOTO 4020,4050
4020 GOSUB 4110: VTAB (10): PRINT "EDIT CASE NUMBER";: INPUT
C%: IF C% < 1 OR C% > NF% THEN 4020
4030 J = C%: GOSUB 4140: GOSUB 4130: IF A$ = CHR$(89) THEN
4020
4040 GOTO 4090
4050 GOSUB 4120: VTAB (10): PRINT "UPPER LIMIT OF VALID CASES";:
INPUT UL: VTAB (12): PRINT "LOWER LIMIT OF VALID CASES";: INPUT
LL: VTAB (14): PRINT "PRINT MISSING<Y/N>";: INPUT B$
4060 GOSUB 4120: VTAB (10): HTAB (10): PRINT "SEARCHING": FOR
J = 1 TO NF%: IF B$ = CHR$(78) AND A(IX,J) = MW THEN 4080
4070 IF A(IX,J) < LL OR A(IX,J) > UL THEN GOSUB 4140
4080 NEXT J: GOSUB 4130: IF A$ = CHR$(89) THEN 4050
4090 IF M% = 0 THEN RETURN
4100 A$ = VCPU$(IX): FOR J = 1 TO NF$:IO(J) = A(IX,J): NEXT
: GOSUB 900: RETURN
4110 HOME : HTAB (10): PRINT "DATAN II EDIT DATA": RETURN
4120 GOSUB 4110: VTAB (3): PRINT "FILE ";F$;: HTAB (20): PRINT
"VARIABLE ";VCPU$(IX): RETURN
4130 GOSUB 4110: VTAB (10): PRINT "FURTHER CORRECTIONS <Y/N>";:
INPUT A$: RETURN

```

```

4140 GOSUB 4120: VTAB (8): PRINT "CASE NUMBER ":J: VTAB (12):
PRINT "CURRENT CODE ":A(I%,J): VTAB (14): HTAB (15): PRINT "EDIT
THIS DATA(Y/N)":; INPUT A#: IF A# = "Y" OR CHR# (68) THEN GOSUB
4110: RETURN
4150 GOSUB 4120: VTAB (8): PRINT "CASE NUMBER ":J: VTAB (12):
PRINT "CURRENT CODE ":A(I%,J): VTAB (14): PRINT "NEW CODE
":; INPUT A(I%,J): GOSUB 4120:MX = 1: RETURN
5000 GOSUB 190: IF IS% < NV% THEN VTAB (10): CALL - 211:
PRINT ".DATA INPUT NOT COMPLETED": GOSUB 220: RETURN
5010 VTAB (6): PRINT "ENTER N OF ADDITIONAL CASES":; INPUT
N2%: IF N2% < = 0 THEN RETURN
5020 IF NFX + N2% > 200 THEN GOSUB 190: VTAB (6): CALL -
211: PRINT ".N OF CASES >200": GOSUB 220: GOTO 5010
5030 IN% = NFX + 1:NFX = NFX + N2%:IS% = 1: GOSUB 760: RETURN
6000 GOSUB 190: VTAB (10): PRINT "NUMBER OF ADDITIONAL VARIABLES":;
INPUT N2%: IF N2% < = 0 THEN RETURN
6010 IF NV% + N2% > 82 THEN GOSUB 190: VTAB (10): CALL -
211: PRINT ".ONLY 82 VARIABLES ARE POSSIBLE": GOSUB 220: GOTO
6000
6020 GOSUB 190: VTAB (10): PRINT "TOTAL NUMBER OF VARIABLES:
":NV% + N2%: GOSUB 230: IF A# = "Y" OR CHR# (68) THEN 6000
6030 GOSUB 6040: GOSUB 6070:IF% = 1: GOSUB 760: RETURN
6040 H% = NV%:NV% = NV% + N2%: FOR U = H% + 1 TO NV%
6050 GOSUB 190: HTAB (1): VTAB (6): PRINT "NAME OF VARIABLE
NO.":;U:; HTAB (25): VTAB (6): INPUT A#: GOSUB 820: IF M% = 1
THEN 6050
6060 V*(U) = A#: NEXT U: RETURN
6070 GOSUB 190: PRINT :PC% = 0: FOR U = H% + 1 TO NV%:PC% =
PC% + 1: IF PC% = 1 THEN GOSUB 190: HTAB (1): VTAB (6): PRINT
"VARIABLE NO.":; HTAB (20): PRINT "NAME":; PRINT
6080 HTAB (5): PRINT U:; HTAB (9): PRINT ".....":; HTAB
(23): PRINT V*(U): IF PC% = 6 THEN PC% = 0: GOSUB 230: IF A#
= CHR# (78) THEN 6120
6090 IF PC% = 6 THEN PC% = 0: GOSUB 230: IF A# = CHR# (79)
THEN 6120
6100 PRINT : NEXT U: GOSUB 230: IF A# = CHR# (78) THEN 6120
6110 RETURN
6120 HTAB (5): VTAB (20): CALL - 958: PRINT "EDIT WHICH LABEL
(CNO.):"; INPUT CX: IF CX < H% + 1 OR CX > NV% THEN 6120
6130 GOSUB 190: HTAB (1): VTAB (4): PRINT "OLD LABEL":; HTAB
(20): VTAB (4): PRINT "NEW LABEL"
6140 HTAB (1): VTAB (6): PRINT V*(CX):; HTAB (20): INPUT A#:
GOSUB 820: IF M% = 1 THEN 6130
6150 V*(CX) = A#: GOTO 6070

```

```

7000 GOSUB 190: VTAB (10): PRINT "DELETE WHICH VARIABLE";:
INPUT A$: GOSUB 260: FOR U = 1 TO NV%: IF V$(U) = A$ THEN CX
= U: GOTO 7020
7010 NEXT U: GOSUB 190: VTAB (10): PRINT "VARIABLE ";A$: PRINT
" DON'T EXIST": GOSUB 220: GOTO 7000
7020 GOSUB 190: VTAB (4): HTAB (6): PRINT "READY TO DELETE
VARIABLE ";V$(CX): VTAB (8): PRINT "SURE ABOUT DELETING VARIABLE
";V$(CX): GOSUB 240
7030 GOSUB 190: VTAB (10): PRINT "DELETING VARIABLE ";V$(CX):N$
= V$(CX) + F2$ + F$:D$ = CHR$ (4): PRINT D$"DELETE ";N$: PRINT
D$"DELETE";VQ$
7040 V$(CX) = V$(NV%):V$(NV%) = CHR$ (32):NV% = NV% - 1:IS'
= IS% - 1: GOSUB 760: RETURN
8000 GOSUB 190: VTAB (6): HTAB (4): PRINT "READY TO DELETE
ENTIRE FILE ";F$: VTAB (10): PRINT "SURE ABOUT DELETING ENTIRE
FILE ";F$: GOSUB 240
8010 FOR U = 1 TO IS% - 1:N$ = V$(U) + F2$ + F$: GOSUB 190:
VTAB (10): PRINT "DELETING ";N$:D$ = CHR$ (4): PRINT D$"DELETE
";N$: NEXT U
8020 GOSUB 190: VTAB (10): PRINT "DELETING ";VQ$:D$ = CHR$
(4): PRINT D$"DELETE ";VQ$: GOSUB 190: VTAB (10): PRINT "ENTIRE
FILE ";F$: PRINT " DELETED": GOSUB 220: CLEAR :LF% = 1: GOTO
40
9000 GOSUB 190: VTAB (10): PRINT "SURE TO COPY DISKETTE";:
INPUT A$: IF A$ < > CHR$ (89) THEN RETURN
9010 GOSUB 190: GOSUB 720: GOSUB 190: VTAB (10): HTAB (10):
PRINT "LOADING COPYPROGRAM":D$ = CHR$ (4): PRINT D$"RUN COPY"
10000 GOSUB 10050:CM% = 16: GOSUB 50000: RETURN
10010 GOSUB 10050: VTAB (3): PRINT "ON THE LEFT SIDE OF THE
ASSIGNING "; PRINT "OPERATOR '=' YOU HAVE TO DEFINE A NEW":
PRINT "VARIABLE NAME."
10020 PRINT "ON THE RIGHT SIDE OF '=' YOU HAVE TO": PRINT "DEFINE
THE VARIABLES AND THE OPERATORS": PRINT "IN THE SAME MANNER
AS IN 'SPSS-COMPUTE'"
10030 PRINT "UP TO 10 VARIABLES ON THE RIGHT SIDE OF": PRINT
" '=' ARE POSSIBLE.": PRINT : HTAB (5): PRINT "<LIST OF POSSIBLE
OPERATORS>": PRINT "ABS ATN COS EXP INT LOG SGN SIN SQR TAN
+ - / * ^ ( )"
10040 PRINT : PRINT "EX.:"; INVERSE : PRINT "NEWVAR = OLD1+
SIN(OLD2)/LOG(OLD3)": NORMAL : PRINT : PRINT "EX.:"; INVERSE
: PRINT "SUM = VAR001 + VAR002 + VAR003": NORMAL : GOSUB 200:
RETURN
10050 HOME : HTAB (10): PRINT "DATAN II COMPUTE": RETURN

```

```

11000 GOSUB 11090: GOSUB 400: VTAB (10): PRINT "ENTER NUMBER
OF VARIABLES TO COUNT";: INPUT N1%: IF N1% > 15 OR N1% < 1 THEN
11000
11010 FOR G = 1 TO N1%: GOSUB 11090: VTAB (8): HTAB (5): PRINT
"VARIABLE ";G;: PRINT " OF VARIABLE LIST": GOSUB 500:VM%(G)
= I%: NEXT G
11020 GOSUB 11090: VTAB (10): PRINT "NUMBER OF CODES TO COUNT";:
INPUT N2%: IF N2% < 1 OR N2% > 9 THEN 11020
11030 GOSUB 11090: VTAB (3): PRINT "DESIRED VARIABLE CODES
TO COUNT": PRINT : FOR G = 1 TO N2%: PRINT "CODE ";G;" ";;
INPUT C(G): PRINT : NEXT G: GOSUB 230: IF A$ < > CHR$(89)
THEN GOTO 11030
11040 GOSUB 800: GOSUB 300: GOSUB 11090: VTAB (10): HTAB (10):
PRINT "COUNTING": FOR J = 1 TO NF%: FOR I = 1 TO N1%:A = A(VM%(I),J):
IF A = MW THEN A(0,J) = MW: GOTO 11080
11050 FOR G = 1 TO N2%: IF A = C(G) THEN A(0,J) = A(0,J) +
1: GOTO 11070
11060 NEXT G
11070 NEXT I
11080 NEXT J: GOSUB 210: GOSUB 950: RETURN
11090 HOME : HTAB (12): PRINT "DATAN II COUNT": RETURN
12000 GOSUB 12310: GOSUB 400: VTAB (5): PRINT "ENTER NAME OF
VARIABLE TO RECODE": GOSUB 500
12010 GOSUB 12310: VTAB (3): HTAB (9): PRINT "SELECT RECODING
OPTION"
12020 VTAB (5): HTAB (7): PRINT "CODE < CONSTANT.....1
": VTAB (7): HTAB (7): PRINT "CODE > CONSTANT.....2 "
VTAB (9): HTAB (7): PRINT "CODE = CONSTANT.....3"
12030 VTAB (11): HTAB (7): PRINT "CODE <> CONSTANT.....4":
VTAB (13): HTAB (7): PRINT "INTERVALS.....5": VTAB
(15): HTAB (7): PRINT "RECODE MISSING.....6": VTAB (17):
HTAB (7): PRINT "SAVE RECODED VARIABLE.....7"
12040 VTAB (23): PRINT "<SAVE VARIABLE TO RETURN TO MENU":
VTAB (19): HTAB (7): INPUT "WHICH....."(H%:
IF A% < 1 OR A% > 7 THEN 12010
12050 ON A% GOSUB 12060,12090,12120,12150,12180,12260,12350:
GOTO 12010
12060 GOSUB 12320: GOSUB 12300: FOR J = 1 TO NF%: IF A(I%,J)
= MW THEN GOTO 12080
12070 IF A(I%,J) < H THEN A(I%,J) = H1
12080 NEXT J: GOSUB 210: RETURN
12090 GOSUB 12320: GOSUB 12300: FOR J = 1 TO NF%: IF A(I%,J)
= MW THEN GOTO 12110
12100 IF A(I%,J) > H THEN A(I%,J) = H1
12110 NEXT J: GOSUB 210: RETURN
12120 GOSUB 12320: GOSUB 12300: FOR J = 1 TO NF%: IF A(I%,J)
= MW THEN GOTO 12140
12130 IF A(I%,J) = H THEN A(I%,J) = H1
12140 NEXT J: GOSUB 210: RETURN

```



```

12150 GOSUB 12320: GOSUB 12300: FOR J = 1 TO NF%: IF A(I%,J)
= MW THEN GOTO 12170
12160 IF A(I%,J) < > R THEN A(I%,J) = H1
12170 NEXT J: GOSUB 210: RETURN
12180 GOSUB 12310: VTAB (10): PRINT "ENTER NUMBER OF INTERVALS":
INPUT N1%: IF N1% < 1 OR N1% > 9 THEN GOTO 12180
12190 FOR G = 1 TO N1%: GOSUB 12310: VTAB (6): PRINT "INTERVAL
NO.":G: VTAB (10): PRINT "ENTER UPPER LIMIT ": INPUT R(G):
VTAB (12): PRINT "ENTER NEW CODE ": INPUT L(G): NEXT
12200 GOSUB 12310: VTAB (3): PRINT "INTERVAL": HTAB (10):
PRINT "UPPER LIMIT": HTAB (30): PRINT "NEW CODE"
12210 PRINT : FOR G = 1 TO N1%: HTAB (4): PRINT G: HTAB (10):
PRINT R(G): HTAB (30): PRINT L(G): PRINT : NEXT : VTAB (13):
HTAB (28): PRINT "OKAY(Y/N)": INPUT A$: IF A$ < > "Y" THEN
RETURN
12220 GOSUB 12300: FOR J = 1 TO NF%: IF A(I%,J) = MW THEN
GOTO 12250
12230 FOR G = 1 TO N1%: IF A(I%,J) < = R(G) THEN A(I%,J) =
L(G): GOTO 12250
12240 NEXT G
12250 NEXT J: GOSUB 210: RETURN
12260 GOSUB 12310: VTAB (10): PRINT "NEW CODE FOR MISSING VALUES":
INPUT H: GOSUB 12310: VTAB (3): PRINT "ATTENTION:THE MISSING
WILL BE VALID": VTAB (5): PRINT "AFTER RECODING."
12270 VTAB (8): PRINT "CURRENT MISSING CODE.....":MW: VTAB
(10): PRINT "NEW CODE FOR THIS.....":H: GOSUB 230: IF
A$ < > CHR$(89) THEN RETURN
12280 GOSUB 12300: FOR J = 1 TO NF%: IF A(I%,J) = MW THEN A(I%,J)
= H
12290 NEXT J: GOSUB 210: RETURN
12300 GOSUB 12310: VTAB (10): HTAB (10): PRINT "RECODING":
RETURN
12310 HOME : HTAB (10): PRINT "< DATAN II RECODE >": RETURN

12320 GOSUB 12310: VTAB (6): PRINT "ENTER VALUE OF CONSTANT":
INPUT H: VTAB (8): PRINT "ENTER NEW VALUE": INPUT H1
12330 GOSUB 12310: VTAB (6): PRINT "CRITERION.....":H:
VTAB (8): PRINT "NEW VALUE.....":H1: GOSUB 230: IF A$ <
> CHR$(89) THEN POP
12340 RETURN
12350 FOR J = 1 TO NF%:A(0,J) = A(I%,J): NEXT : GOSUB 800:
GOSUB 950:VCPU$(IX) = A$: POP : RETURN

```

```

13000 GOSUB 13010:CM% = 22: GOSUB 50000: RETURN
13010 HOME : HTAB (10): PRINT "< DATAN II IF >": RETURN
13020 GOSUB 13010: PRINT : PRINT "DATAN II IF ASSIGNS A SPECIFIED
VALUE": PRINT "TO A VARIABLE IF THIS OR AN OTHER": PRINT "VARIABLE
FULLFIL A USER SPECIFIED": PRINT "CONDITION."
13030 PRINT "THE FIRST STATEMENT MUST BE AN 'IF'.": PRINT "AFTER
'IF' YOU HAVE TO DEFINE THE": PRINT "CONDITION.THESE STATEMENTS
MUST BE": PRINT "FOLLOWED BY A 'THEN' STATEMENT."
13040 PRINT "'THEN' MUST BE FOLLOWED BY THE NAME OF": PRINT
"THE RESULT-VARIABLE.THIS NAME MUST": PRINT "BE FOLLOWED BY
'='AND THE ASSIGNING": PRINT "OPERATION."
13050 PRINT : INVERSE : PRINT "ATTENTION:IF YOU USE AN EXISTING
NAME": PRINT "AFTER '=' THEN THE ORIGINAL DATA OF": PRINT "THIS
VARIABLE WILL BE DELETED.": NORMAL
13060 PRINT : HTAB (10): PRINT "<LIST OF POSSIBLE OPERATORS>":
FOR I = 1 TO 22: PRINT OP$(I):" "": NEXT I: PRINT : GOSUB 200:
GOSUB 13010: PRINT : PRINT "EX.": INVERSE : PRINT "IF AGE
18 AND MEMBER=1 THEN INDEX=1": NORMAL
13070 PRINT "'ELSE'-OPERATION IS IMPOSSIBLE.": PRINT "CASES
WHICH DON'T FULLFIL THE CONDITION": PRINT "WILL BE NOT PROCESSED.THEIR
VALUES": PRINT "REMAINING UNCHANGED.IF THE RESULTING": PRINT
"VARIABLE IS A NEW VARIABLE THESE CASES"
13080 PRINT "WILL BE SET TO ZERO.": PRINT "THE SYNTAX OF 'DATAN
II IF' IS NOT ": PRINT "EQUAL TO THE 'SPSS-IF' BUT VERY SIMILAR.":
GOSUB 200: RETURN
14000 GOSUB 14060: GOSUB 400: GOSUB 500: GOSUB 800: GOSUB 14060:
VTAB (10): HTAB (10): PRINT "COMPUTING":UL = 0:LL = 0:MC% =
0: FOR J = 1 TO NF%:H = A(I%,J): IF H = MW THEN MC% = MC% +
1: GOTO 14020
14010 UL = UL + H:LL = LL + H * H
14020 NEXT :H = NF% - MC%:I = UL : H:G = LL / H:UL = (UL /
H) ^ 2:G = G - UL: IF G < = 0 THEN GOSUB 14060: VTAB (10):
CALL - 211: PRINT ".STD.DEV. IS ZERO.": VTAB (12): PRINT "PROCEDURE
ABORTED.": GOSUB 200: RETURN
14030 G = SQR (G): FOR J = 1 TO NF%: IF A(I%,J) = MW THEN A(0,J)
= MW: GOTO 14050
14040 A(0,J) = (A(I%,J) - I) / G
14050 NEXT : GOSUB 950: RETURN
14060 HOME : HTAB (5): PRINT "< DATAN II Z-TRANSFORMATION >":
RETURN
15000 CLEAR : GOSUB 210: GOTO 40
16000 PC% = 1: GOSUB 16020: VTAB 3: FOR I = 1 TO NV%: HTAB (PC%):
PRINT MID$(V$(I),1,6): IF PEEK (37) > 21 THEN PC% = PC% +
8: VTAB 3
16010 NEXT : GOSUB 200: RETURN
16020 LX = (19 - LEN (F$)) / 2: HOME : INVERSE : FOR U = 1
TO LX: PRINT " "": NEXT : PRINT "VARIABLECATALOG FILE ":F$,:
FOR U = 1 TO LX: PRINT " "": NEXT : NORMAL : RETURN

```

```

17000 GOSUB 280: PRINT "RETURN TO MENU":D$ = CHR$(4): PRINT
D$:"RUN REP-TEST MENU"
50000 GOSUB 50360: VTAB(10): PRINT "INITIALIZING":PA = PEEK
(103) + PEEK(104) * 256: FOR AD = PA TO PA + 4000: IF PEEK
(AD) < > 58 THEN NEXT
50010 IF PEEK(AD + 1) = 58 THEN 50030
50020 NEXT : GOTO 60070
50030 GOSUB 300:H% = 0:C% = 0:C$ = "":B$ = C$:N2% = 0:RC% =
0:LC% = 0:N1% = 0:PP% = 0: GOSUB 50360: VTAB(5): HTAB(7):
PRINT "ENTER COMPUTING FORMULA": VTAB(10): INPUT "":A$: IF
A$ = "" THEN RETURN
50040 IF A$ = "H" AND CM% = 16 THEN GOSUB 10010: GOTO 50030
50050 IF A$ = "H" AND CM% = 22 THEN GOSUB 13020: GOTO 50030
50060 FOR G = 1 TO LEN(A$): IF MID$(A$,G,1) < > CHR$(
(32)) THEN B$ = B$ + MID$(A$,G,1)
50070 NEXT :C$ = B$: IF CM% = 16 THEN 50230
50080 IF LEFT$(B$,2) < > "IF" THEN 50430
50090 B$ = MID$(B$,3, LEN(B$)):PP% = PP% + 1: POKE AD,173:
GOTO 50190
50100 GOSUB 50360: VTAB(3): HTAB(10): PRINT "COMPILING":
VTAB(5): PRINT C$: LIST 20: IF LEN(B$) = 0 THEN 50470
50110 IF N1% > 10 THEN 50390
50120 IF PP% > 200 THEN 50400
50130 IF ASC(B$) = 46 THEN GOSUB 50270:B$ = MID$(B$,2,
LEN(B$)): IF ASC(B$) < 48 OR ASC(B$) > 57 THEN 50370
50140 IF ASC(B$) > 47 AND ASC(B$) < 58 THEN 50260
50150 IF ASC(B$) = 40 THEN LC% = LC% + 1: GOTO 50260
50160 IF ASC(B$) = 41 THEN RC% = RC% + 1: GOTO 50260
50170 FOR SP = 1 TO 4:S$ = LEFT$(B$,SP): FOR O = 1 TO CM%:
IF S$ = OP$(O) THEN GOSUB 50280: GOTO 50100
50180 NEXT O: NEXT SP
50190 FOR Z = 7 TO 1 STEP - 1: FOR I = 1 TO NV%: IF LEFT$(
B$,Z) < > V$(I) OR Z > LEN(B$) THEN NEXT I: NEXT Z: GOTO
50210
50200 B$ = MID$(B$,Z + 1, LEN(B$) - Z):A$ = V$(I): GOSUB
530:N1% = N1% + 1:VM$(N1%) = I$: GOSUB 50330: GOTO 50100
50210 IF N2% = 1 THEN GOTO 50440
50220 IF CM% = 22 AND PP% = 1 THEN 50410
50230 FOR Z = 2 TO 7: IF MID$(B$,Z,1) < > CHR$(61) THEN
NEXT Z: GOTO 50410
50240 Z = Z - 1:A$ = MID$(B$,1,Z): GOSUB 820: IF M% = 1 THEN
GOSUB 50320: GOTO 50030
50250 N2% = 1:VCP$(0) = A$:I% = 0:B$ = MID$(B$,Z + 1, LEN
(B$) - Z): GOSUB 50340: GOTO 50100
50260 GOSUB 50270:B$ = MID$(B$,2, LEN(B$)): GOTO 50100
50270 POKE AD + PP%, ASC(B$):PP% = PP% + 1: RETURN
50280 IF O < 11 AND ASC(MID$(B$,SP + 1,1)) < > 40 THEN
50380
50290 POKE AD + PP%,OC$(O):PP% = PP% + 1: IF O = 20 THEN C%
= C% + 1:A% = - 1
50300 C(1) = LEN(B$):C(2) = LEN(OP$(O)): IF C(1) - C(2)
= 0 THEN 50410
50310 B$ = RIGHT$(B$, LEN(B$) - SP): RETURN

```

```

50320 FOR I = AD TO AD + PP%: POKE I,58: NEXT I: RETURN
50330 IF AX = - 1 THEN AX = 0:HX = IX
50340 POKE AD + PP%,65:PP% = PP% + 1: POKE AD + PP%,40:PP%
= PP% + 1: POKE AD + PP%,ASC (STR$(IX))
50350 PP% = PP% + 1: POKE AD + PP%,44:PP% = PP% + 1: POKE AD
+ PP%,74:PP% = PP% + 1: POKE AD + PP%,41:PP% = PP% + 1: RETURN

50360 HOME : HTAB (10): PRINT "DATAN II INTERPRETER": RETURN

50370 GOSUB 50450: PRINT ".INVALID CHARACTER AFTER ' '": GOTO
50460
50380 GOSUB 50450: PRINT ".PARENTHESES MISSING AFTER '":OP$(0):" '":
GOTO 50460
50390 GOSUB 50450: PRINT ".MORE THAN 10 VARIABLES": GOTO 50460
50400 GOSUB 50450: PRINT ".COMPILED FORMULA TOO LONG": GOTO
50460
50410 GOSUB 50450: PRINT ".INVALID OPERATOR OR VARIABLENAME":
GOTO 50460
50420 GOSUB 50450: PRINT ".N OF '(' NOT EQUAL N OF ')'": GOTO
50460
50430 GOSUB 50450: PRINT ".IF/ OPERATOR MISSING": GOSUB 200:
GOTO 50030
50440 GOSUB 50450: PRINT ".MORE THAN ONE RESULT-VARIABLE":
GOTO 50460
50450 GOSUB 50360: VTAB (6): PRINT C$: VTAB (10): CALL - 211:
RETURN
50460 GOSUB 50320: GOSUB 200: GOTO 50030
50470 IF LC% < > RC% THEN GOTO 50420
50480 IF CM% = 22 AND C% < > 1 THEN 50410
50490 GOSUB 50360: VTAB (10): HTAB (10): PRINT "COMPUTING":
FOR J = 1 TO NF%: VTAB (12): HTAB (10): PRINT "CASE";J: FOR
U = 1 TO N1%: IF A(VM$(U),J) = MW THEN A(H%,J) = MW: GOTO 50510
50500 NEXT U: GOSUB 20
50510 NEXT J:A$ = VCPU$(H%): FOR J = 1 TO NF%:IO(J) = A(H%,J):
NEXT J: GOSUB 900: GOSUB 50320: IF N2% = 1 THEN GOSUB 810:
GOSUB 760
50520 RETURN

```


IV. Fragebogen zur Untersuchung

"Ausländische Studenten der RUB"

Teil I

(wenn sie diesen Teil nicht beantworten möchten, machen Sie bitte bei Teil II weiter)

1. Geschlecht: männlich 0 weiblich 0

2. Alter: Jahre

3. Herkunftsland:

4.1. Studienfächer: /

4.2. Ich habe vor meinem Studium in der Bundesrepublik Deutschland noch nicht an einer Universität studiert

ja 0 nein 0

5. Semesterzahl: 1. Sem. 0 2. - 5. Sem. 0

6. - 10. Sem. 0 mehr als 10 Sem. 0

6. Ich bin seit . . . Jahren in der Bundesrepublik Deutschland

7. Familienstand: ledig 0 verheiratet 0 sonstiges 0

8. angestrebter Studienabschluß (zum Beispiel Magister, Diplom etc.):

.

9. Ich habe schon einmal an einer Prüfung (z.B. Zwischenprüfung, Vordiplom, Diplom) während meines Studiums in der Bundesrepublik Deutschland teilgenommen

ja 0 nein 0

wenn ja: ich habe die letzte von mir abgelegte Prüfung bestanden:

ja 0 nein 0

ich habe schon einmal eine Prüfung im Rahmen des Studiums nicht bestanden:

ja 0 nein 0

10. Ich habe mein Studienfach schon einmal gewechselt

ja 0 nein 0

wenn ja: (bitte Studienfächer angeben)

vorher habe ich studiert,

dann habe ich studiert,

jetzt studiere ich

11. Meine Sprachkenntnisse in Deutsch habe ich in erster Linie auf folgende Weise erworben:

im Heimatland 0

in der Bundesrepublik
Deutschland 0

auf sonstige Weise

12. Ich finanziere mein Studium hauptsächlich auf folgende Weise
(bitte nur eine Antwort)

selbst	0	Eltern oder Verwandte	0	deutsches Stipendium	0
ausländisches Stipendium	0	auf sonstige Weise			

Teil II

13. Ich habe/hatte in meinem Studium folgende Schwierigkeiten:
(mehrfach ankreuzen möglich)

Dies war/ist

- | | |
|--|---|
| 13.1. wegen des Notendurchschnitts meines Schulabschlußzeugnisses | 0 |
| 13.2. bei der Deutschprüfung (falls sie erforderlich war) | 0 |
| 13.3. bei der Feststellungsprüfung (falls sie erforderlich war) | 0 |
| 13.4. weil ich <u>vor</u> Zulassung zum Studium nicht genügend über das Studium und dessen Voraussetzungen (Deutschkenntnisse, Visa etc.) informiert wurde | 0 |
| 13.5. weil ich über den Ablauf des Studiums auch während des Studiums | |
| falsch | 0 |
| widersprüchlich | 0 |
| nicht ausreichend | 0 |
| informiert wurde | |
| 13.6. weil mir zu Beginn des Studiums wichtige Informationen, die zum Leben in Deutschland wichtig sind, fehlten | 0 |
| 13.7. weil ich Schwierigkeiten mit den Behörden in Deutschland habe | 0 |
| 13.8. weil es mir schwer fällt den Stoff der Vorlesung zu begreifen | 0 |
| 13.9. weil ich finanzielle Probleme habe | 0 |
| 13.10. weil ich keinen Platz habe, an dem ich in Ruhe arbeiten kann | 0 |
| 13.11. weil ich Schwierigkeiten mit meiner Unterbringung (Wohnung) habe | 0 |
| 13.12. weil ich Schwierigkeiten beim Lernen habe | 0 |
| 13.13. (sonstiges) | |

14. Die Gründe für diese Schwierigkeiten sind vor allem die folgenden:
(Mehrfachnennung möglich)

- | | |
|---|---|
| 14.1. Meine Schulausbildung war schlechter als die der deutschen Kommilitonen | 0 |
| 14.2. Ich habe Schwierigkeiten mit der deutschen Sprache | 0 |
| 14.3. Mir fehlen/fehlten hinreichende Informationen über den Studienablauf | 0 |
| 14.4. Wegen anderer (privater/persönlicher) Probleme konnte ich zeitweise nicht studieren | 0 |
| 14.5. Wegen meiner schlechten Wohnverhältnisse kann ich nicht richtig studieren | 0 |
| 14.6. (sonstiges) | |
| 15.1. Ich kenne das "Studienbüro" | |

ja	0	nein	0
----	---	------	---

- | | | | | |
|--|----|---|------|---|
| 15.2. Ich kenne besondere Einrichtungen (Tutorien etc.) für ausl. Studenten an unserer Universität | ja | 0 | nein | 0 |
| 15.3. Ich fände es gut, wenn es besondere Einrichtungen (Tutorien etc.) für ausl. Studenten gäbe | ja | 0 | nein | 0 |

16. Wenn ich Schwierigkeiten habe, wende ich mich vor allem an:
(Mehrfachnennung möglich)

- | | |
|--|---|
| 16.1. das "Studienbüro" | 0 |
| 16.2. Landsleute | 0 |
| 16.3. deutsche Kommilitonen | 0 |
| 16.4. an die Studienberatung unserer Abteilung | 0 |
| 16.5. ich wende mich an niemanden, ich arbeite allein | 0 |
| 16.6. die Fachschaft | 0 |
| 16.7. Professoren / Assistenten | 0 |
| 16.8. das Akademische Auslandsamt | 0 |
| 16.9. die ESG / KSG | 0 |
| 16.10. den AstA | 0 |
| 16.11. (sonstiges)
(bitte nennen Sie ein Stichwort) | |

7. Fühlen Sie sich manchmal an der Universität und / oder auch außerhalb der Universität diskriminiert, weil sie Ausländer sind?

ja 0 nein 0

wenn ja: bei welchen Gelegenheiten? (Stichworte genügen)

.....

8. Ich möchte meine deutschen Sprachkenntnisse verbessern

ja 0 nein 0

9. Ich weiß ~~nicht~~, wie ich meine deutschen Sprachkenntnisse verbessern kann

ja 0 nein 0

- | | trifft zu | trifft weniger zu | trifft zur Hälfte zu | trifft überwiegend nicht zu | trifft gar nicht zu |
|--|-----------|-------------------|----------------------|-----------------------------|---------------------|
| 10. Ich arbeite ständig mit mehreren deutschen Kommilitonen zusammen | +2
0 | +1
0 | +0
0 | -1
0 | -2
0 |
| 11. Ich arbeite ständig mit mehreren Kommilitonen aus meinem Heimatland zusammen | 0 | 0 | 0 | 0 | 0 |
| 12. Ich arbeite oft alleine | 0 | 0 | 0 | 0 | 0 |
| 13. Ich habe guten privaten Kontakt zu Deutschen | 0 | 0 | 0 | 0 | 0 |
| 14. Ich fühle mich wohl in Deutschland | 0 | 0 | 0 | 0 | 0 |
| 15. Ich habe guten privaten Kontakt zu meinen Landsleuten | 0 | 0 | 0 | 0 | 0 |
| 16. Ich spreche gut Deutsch | 0 | 0 | 0 | 0 | 0 |
| 17. Ich denke, Sie haben mit diesem Fragebogen meine Probleme nicht hinreichend erfaßt | ja 0 | nein 0 | | | |
| 18. Ich möchte Sie deshalb auf Folgendes hinweisen (freien Raum unten benutzen) | | | | | |

V. Codeplan

Frage Nr.	Spalten Anschlüsse	Inhalt	Name	Wertebereich	Werte	Code- rung	missing value
-	1-2	lfd. Nr.	LN	01-75	-	-	-
-	3	Grund f. Nicht-Ausfüll.	V1	0-2, 9	kenne AG nicht Angst sonstiges	0 1 2	9
1	4	Geschlecht	V2	0,1,9	männlich weiblich	0 1	9
2	5-6	Alter	V3	18- 99	-	-	99
3	7	Herkunftsland	V4	0-6, 9	Afrika Amerika Nah-Ost übr. Asien EG Türkei übr. Europa	0 1 2 3 4 5 6	9
4.1	8 9	1. Studienfach 2. Studienfach	V5 V6	0-4, 9	Ing.-wissensch. Philologie sonst. C.-wiss. Naturwiss. Medizin	0 1 2 3 4	9
4.2	10	Vorher noch nicht stud.	V7	0, 1, 9	nein ja	0 1	9
5	11	Semesterzahl	V8	0-3, 9	1. Sem. 2.-5. Sem. 6.-10. Sem.	0 1 2	9

Frage Nr.	Spalten Anschlä- ge	Inhalt	Name	Wertebereich	Werte	Codierung	missing value
6	12-13	wieviel Jahre in BRD?	V9	Ø1- ,99	-	-	99
7	14	Familienstand	V1Ø	Ø-2, 9	ledig verheiratet sonstiges	Ø 1 2	9
8	15	angestrebter Stud.absch.	V11	Ø-5, 9	Diplom Magister Promotion Staatsex. (o. Lehr.) Lehramtsprüfung sonstige	Ø 1 2 3 4 5	9
9	16-18	an Prüfung teilgenomm.? wenn ja, letzte Prüf bestanden? schon einm. u. bestand?	V12 V13 } V14 }	Ø, 1, 9 Ø, 1, 8-9	nein ja nein ja betrifft nicht	Ø 1 Ø 1 8	9 9
10	19-22	Studienfachwechsel ? wenn ja, vorher studiert dann studiert jetzt studiert	V15 V16 } V17 } V18 }	Ø, 1, 9 Ø-4, 8-9	nein ja Ing.-wissensch. Philologie sonst. G.-wiss. Naturwissensch. Medizin betrifft nicht	Ø 1 Ø 1 2 3 4 8	9 9
11	23	dt. Sprachkenntnisse	V19	Ø-2, 9	Heimatland Bundesrepublik sonstiges	Ø 1 2	9

Frage Nr.	Spalten Anschl- ge	I n h a l t	Name	Wertebereich	W e r t e	Code- rung	missing value
12	24	Finanzierung	V20	0-4, 9	selbst Eltern dt. Stipendium ausl. Stipendium sonstiges	0 1 2 3 4	9
13	25-39	Schwierigk. im Studium	V21-V35	0, 1	nein ja	0 1	-
14	40-45	Gründe dafür	V36-V41	0, 1, 2	nein ja betrifft nicht	0 1 2	-
15	46-48	Kenntnisse v. Ausl. Einr.	V42-V44	0, 1, 9	nein ja	0 1	9
16	49-59	Anlaufstelle b. Schwierigkeiten	V45-V55	0, 1	nein ja	0 1	-
17	60-61	Diskrim.-wahrnehmung wenn ja, welche Gelegenheiten?	V56 V57	0, 1, 9 0-8, 9	nein ja betrifft nicht allg. Atmosphäre stud. Arbeitsgen. Assist., Dozenten Uni-Verwaltung Kontaktmögl. Geschäfte, Gaststätt Wohnungssuche Behörden außerh. Uni	0 1 2 3 4 5 6 7	9 9

Frage Nr.	Spalten Anschl- ge	I n h a l t	Name	Werte- reich	W e r t e	Codie- rung	missing value
18	62	Verbess. Sprachkenntn.?	V58	Ø, 1, 9	nein ja	Ø 1	9
19	63	gewußt, wie	V59	Ø, 1, 9	nein ja	Ø 1	9
20 21 22 23 24 25 26	64 65 66 67 68 69 70	Zs.-arbeit m. Dtschen Zs.-arbeit m. Landsl. arbeite allein priv. Kontakt zu Dt. Wohlfühlen in Deutschl. priv. Kont. zu Landsl. ich spreche gut Deutsch	V60 V61 V62 V63 V64 V65 V66	Ø-4, 9	trifft gar n. zu überwiegend nicht zur Hälfte trifft weniger zu trifft zu	Ø 1 2 3 4	9
27	71	Meine Probl. erfaßt	V67	Ø, 1, 9	nein ja	Ø 1	9

Erklärung

Ich erkläre, daß diese Dissertation weder in der gegenwärtigen noch in einer anderen Fassung dieser oder einen anderen Fakultät vorgelegen hat.

Ich versichere, daß ich die Dissertation selbständig angefertigt habe, daß alle Hilfen und Hilfsmittel angegeben, insbesondere die wörtlich oder dem Sinne nach anderen Veröffentlichungen entnommenen Stellen kenntlich gemacht wurden.

Bochum, den 2. Februar 1983

R. Prust

Lebenslauf

24.5.1952	geboren in Essen als Tochter von Rudolf Prust und seiner Ehefrau Luise, geb. Marsch
Frühjahr 1958	Einschulung in die Tiegelschule in Essen
Januar 1959	Wechsel zur Gerswidaschule in Essen
Frühjahr 1959	Übergang zur Maria-Wächtler-Schule, neusprachliches Mädchen-Gymnasium in Essen
10. Juni 1970	Erfolgreiches Absolvieren der Reifeprüfung
1. Oktober 1970	Immatrikulation an der Ruhr-Universität Bochum
20. Dezember 1978	Verleihung des Grades "Diplom-Mathematiker" durch die Abteilung für Mathematik
1. Januar 1979 bis 31. Dezember 1982	Beschäftigung als Wissenschaftliche Hilfskraft an der Abteilung für Sozialwissenschaft, Sektion Methodenlehre und Statistik

